LOCSU

Audit, Evaluate, Improve: Practical Statistics for Optometrists

by David Knight

December 2025

Version 1

Audit, Evaluate, Improve: Practical Statistics for Optometrists

LOCSU

As optometrists, we're constantly gathering data, from patient history and symptoms to refractive status, visual acuity, and frame measurements! This primer - along with the accompanying demonstration spreadsheet - will equip you with the fundamental statistical concepts needed to start to make sense of that data, helping you to conduct meaningful audits and service evaluations.

Why is this important? Because correct statistical arguments remove the guesswork and provide a robust, evidence-based way to measure and improve clinical services and make informed business decisions.

Don't worry, we're not aiming to turn you into a biostatistician. This is simply a "jumping-off point" to help you ask the right questions and interpret your findings. The first part of this article assumes you've already collected your dataset and are staring at a spreadsheet, thinking, "Now what?" We'll tackle the more complex subject of inferential statistics in the second part, but you'll find that for many audits, you don't even need it.

Understanding the Type of Your Data

Now you have your data, the first question to ask is what kind of data have I collected? There are several broad data types, and it's important to understand which data type you are working with as it will affect how you analyse it. For example, average age makes sense, but average eye colour doesn't!

Main Data Types

Categorical *(qualitative): Describes qualities or characteristics.*

- **Nominal** – Categories with no inherent order (e.g. *eye colour*, *glaucoma subtype*)
- **Binary** – A special type of nominal data where there are only two categories (e.g. was a referral made? (*Yes*/*No*), or birth gender assignment)
- **Ordinal** – Categories with meaningful rank or order (e.g. *dry eye severity*, *patient satisfaction*)

Numerical *(quantitative): Represents quantities.*

- **Discrete** – Discrete, separate, whole numbers, generally counts (e.g. number of *patients*, *referrals*, or *dispenses*)
- **Continuous** – Values can be any number, not just whole numbers. Generally measurements (e.g. *IOP*, *contact lens diameter*, *corneal thickness*, *axial length*, *examination time*)

Once you know the type of data you are working with, you can now start to analyse it, as each data type needs to be handled differently. Average eye colour doesn't make sense, but the frequency of each eye colour may be more helpful to determine.

However, this isn't the end of the story, as we also need to understand the general shape of the data.

Making Sense of your Data: Introduction to Descriptive Statistics

Descriptive statistics help us summarise and describe the main features of a dataset. They provide a simple summary about the sample and the measures to help you understand general characteristics of the data you are analysing.

For example, you can use descriptive statistics to find the average age of your patients, or the most common lens coating.

The way you analyse data depends on the data type:

Nominal/Binary Data

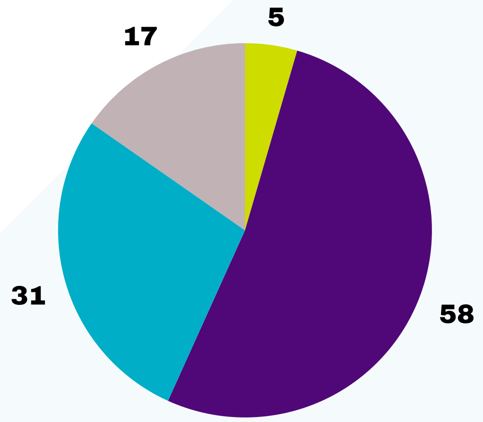
Frequencies or counts can help understand which categories are more, or less, common (e.g. number of *green*, *brown*, *blue* and *hazel* eyes - shown in the two examples on the *right*).

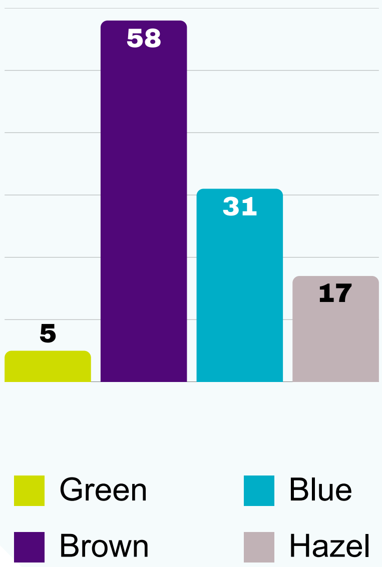
This can then be represented as percentages, a ratio, or visually in the form of a **pie chart** or **bar chart**.

When dealing with many categories (e.g., more than *six*), grouping them can help visualise overall patterns.

The key purpose of a visualisation is to help your audience understand the story in your data, so the level of detail should be appropriate for the message.

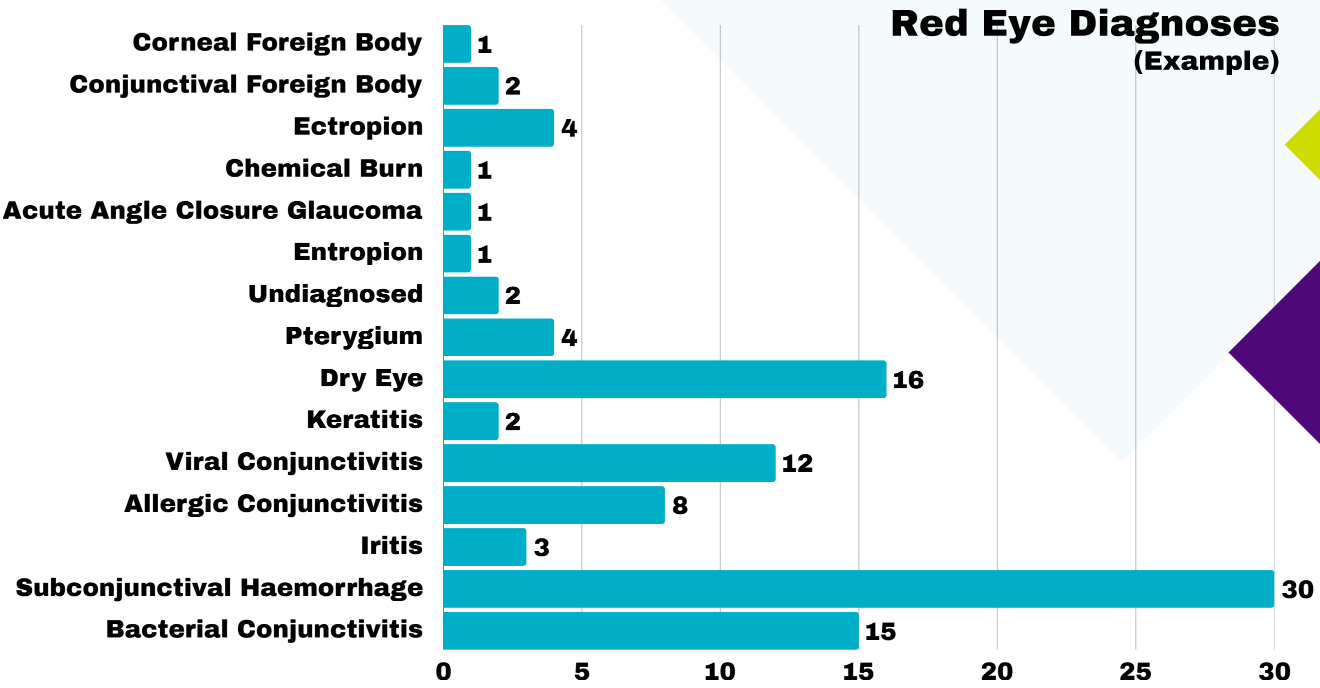
Continued on next page.

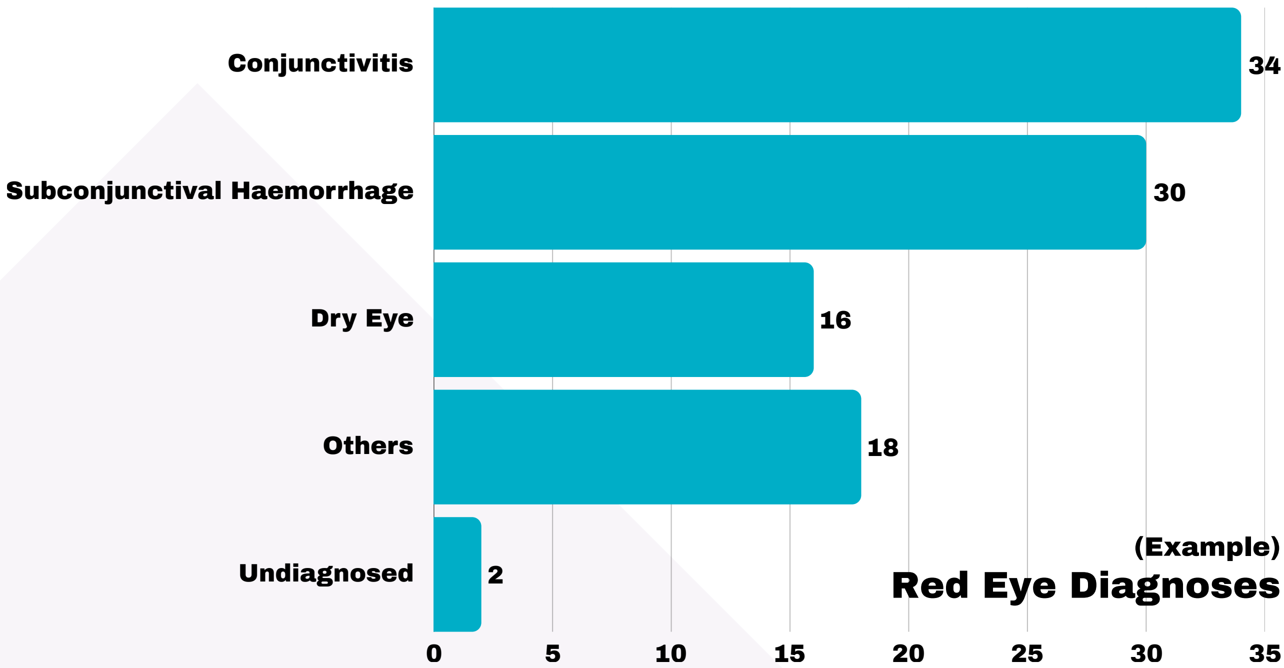




Be mindful of the potential for bias when altering how you group data. In the example **below** concerning *red eye diagnoses*, all conjunctivitis cases were combined. While this offers a simple, high-level view, it can obscure important details.

For instance, chlamydial conjunctivitis requires different management than other forms. If your analysis is about managing red-eye conditions, a more specific classification would therefore be necessary to avoid misrepresenting the data.

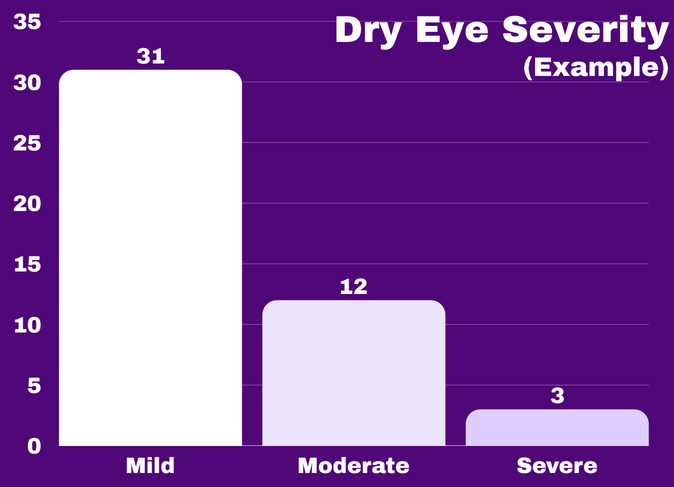




Ordinal Data

Similarly to nominal data, you can also summarise ordinal data with percentages, ratios or proportions, and frequencies or counts.

Similarly, pie charts and bar charts can be used, but as the order of categories is important, **bar charts** are generally better used to visualise ordinal data.

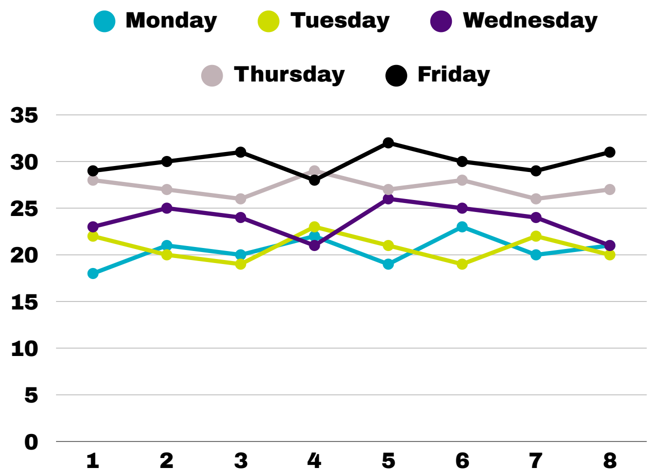


Discrete Data

Like nominal and ordinal data, you can count the frequency of each value (e.g. how many patients were seen on each day).

You can calculate the mean, median and mode, though median and mode are often more representative since the average may not be a possible value (e.g. an average of 1.8 contact lenses ordered per hour).

They are best visualised using bar charts, histograms, scatter plots and **line plots**.



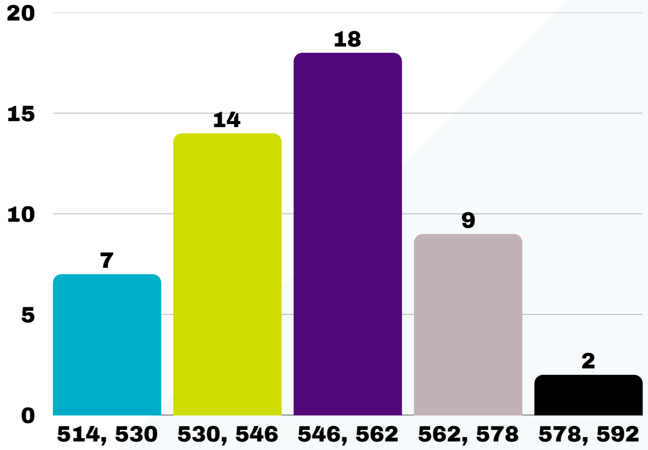
Patients seen on Clinic Days
(Example)

Continuous Data

You can calculate a full set of descriptive statistics for continuous data, including central tendency, variability, and shape.

For visualisations, you will often need to group the continuous data into **categories** to make it easy to interpret.

Histograms are the most common way to visualise this data. Box plots and line graphs can also be helpful.



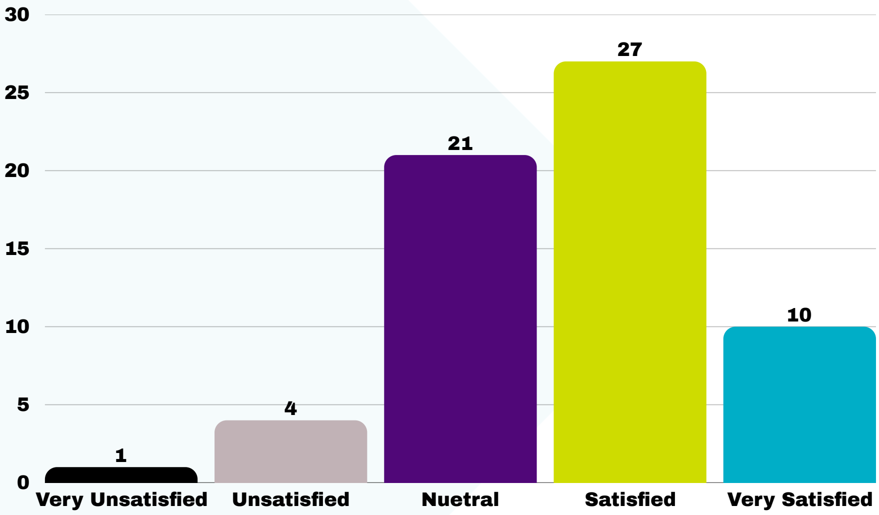
Central Corneal Thickness
(Example)

Calculating Descriptive Statistics

When we think about datasets, the most common shape we think about is the classic “bell-shaped” curve, otherwise known as the normal distribution.

Whilst a lot of data does generally fit this curve, this isn’t always the case.

For example, dry eye severity is mostly mild, so the graph will be significantly "skewed" to the left.



Similarly, the graph of patient satisfaction (*above*) is slightly skewed to the right, as more patients were satisfied than dissatisfied.

Understanding if the dataset is like the normal distribution helps us to know which statistical tests we can perform.

We can visualise these patterns with graphs, but it can be helpful to calculate the values that represent these properties so we can compare our results.

The key figures that help us to understand most datasets are:

- **The central tendency**, which describes the “middle” or “average” of the dataset (mean, median, and mode)
- **The variability or dispersion**, which describe how spread out the data is (range, variance, standard deviation)
- **The shape of the data**, which describe the general form of a distribution (skewness, kurtosis)

The following **Excel formulas** assume that the dataset you are working with is in column A and between rows 2 and 100. You will need to change this for your data.

Measure	Formula	Description
Mean	=AVERAGE(A2:A100)	The sum of all values divided by the number of values. The most common measure of central tendency but can be affected by extreme values (outliers)
Median	=MEDIAN(A2:A100)	The middle value in an ordered dataset. It is less affected by extreme values (outliers) than the mean, which can make it more representative.
Mode	=MODE.SNGL(A2:A100)	The value that appears most frequently in a dataset. A dataset can have no mode, one mode, or multiple modes.
Range	=MAX(A2:A100)-MIN(A2:A100)	The difference between the highest and lowest values in the dataset. It gives a quick, simple sense of the data's spread. Heavily affected by outliers.
Variance	=VAR.S(A2:A100)	The average of the squared differences from the mean. It gives an idea of how much individual data points vary from the average.
Standard Deviation	=STDEV.S(A2:A100)	The square root of the variance. A low SD means data points are clustered near the mean, whilst a high SD means they are more spread out (dispersed)
Skew	=SKEW(A2:A100)	Measures the asymmetry of the data distribution. A negative value means the "tail" is longer on the left (skewed left), whilst a positive value means the "tail" is on the right (skewed right). A value of zero indicates a symmetrical distribution.
Kurtosis	=KURT(A2:A100)	Measures the "tailedness" of the distribution, or how many outliers are present compared to a normal distribution. A higher value means more outliers.

Telling the Story of your Data: Interpreting your Findings

So you've calculated descriptive statistics, you have the building blocks for your service evaluation or audit. Now comes the important bit: interpreting these numbers to tell a meaningful story. These findings help you move from simply describing your data to understanding its real-world implications for patient care and practice performance.

Central Tendency

In a normal distribution, the mean, median, and mode (for discrete data) will be nearly the same. If they are not, your dataset is likely to be skewed or contain outliers.

Case Study: Patient Age. If the mean age of your patients is significantly higher than the median, it suggests that a few much older patients are pulling the average up, which could influence your choice of frame stock.
Case Study: Visual Acuity. In a study of visual acuity outcomes after cataract surgery, comparing the mean and median can reveal if a few patients with very poor results are skewing the overall "average" outcome.

Variability

Variability measures how spread out your data is. A low standard deviation means data points are clustered tightly around the mean, while a high standard deviation indicates a wider spread.

Case Study: Examination Time. A low standard deviation for examination time means your optometrists are very consistent. A high standard deviation means there is significant variability, and it would be worth investigating why some appointments are much longer or shorter than others.
Case Study: Intraocular Pressure (IOP). When measuring IOP in a patient over time, a low standard deviation indicates stable pressure, while a high standard deviation suggests fluctuations that could be concerning for glaucoma management.

Outliers

Outliers are data points that differ significantly from other observations. Watch for outliers; they tell stories! They can be genuine, but they can also be the result of a transcription error or an unusually extreme event.

Case Study: Patient Satisfaction. An outlier in patient satisfaction scores (e.g., a single low score of 1 while all other scores are 6–8) could be a genuinely poor patient experience, or it could be due to a data entry error. It is important to investigate.
Case Study: Referral Rates. An optometrist with a significantly higher referral rate for a specific condition might be an outlier. This could mean they are either correctly identifying more cases or have a lower referral threshold that could warrant discussion.

Visualising the Data

Graphs and charts bring your data to life and make it accessible to everyone, including non-statisticians.

- **Before/After Comparison**: A bar chart can show the number of patients in each satisfaction category before and after a new patient communication strategy. A histogram of visual acuity measurements can show if the distribution has shifted toward better results after a new treatment.
- **Operational Insights:** A bar chart of weekly patient volume can make it obvious that Mondays and Fridays are busiest, which could inform your staffing rota. A box plot of examination times can quickly highlight outliers that might require follow-up.

Putting it All Together

By combining your understanding of data types, descriptive statistics, and visualisations, you can now construct and communicate a clear narrative about your findings.

For example:

- "Our audit showed that patient satisfaction is high overall, with a median score of 8. However, the data was slightly skewed to the left, and a bar chart showed we had a small but noticeable group of patients with a very low satisfaction score. Further investigation revealed these were linked to long waiting times."
- "The average CCT in our sample was 545 μm with a standard deviation of 30 μm, which is consistent with the normal population. A histogram showed a clear bell-shaped curve, indicating our patient base is representative and does not contain any obvious outliers that could invalidate our findings."

Making Inferences from your Data: An Introduction to Inferential Statistics

Descriptive statistics help you summarise your data, and for most audits or service evaluations, that's usually enough. But sometimes, you'll want to go a step further: can we be confident that what we found in our sample applies to the wider population of patients?

That's where inferential statistics come in.

They allow you to use a sample to make an educated guess ("inference") about a larger group. For example, you might collect waiting time data from 50 patients, but you want to know whether this represents all patients attending your clinic.

Inferential statistics is an enormous subject that I cannot hope to summarise in an introductory primer, so I have significantly pared down this section. If you want to learn more about this subject, I'd recommend you check out the resources at the end of this article. With that said, let's dive in!

Step One:
Frame Your Question as a Hypothesis

LOCSU

Every inferential test begins with a pair of competing statements:

- **Null Hypothesis (H_0):** Assumes there is no effect, difference, or relationship. Example: "There is no difference in average waiting times before and after introducing the new booking system."
- **Alternative Hypothesis (H_1):** Suggests that there is a difference or relationship. Example: "The new booking system reduces average waiting times."

After testing, you will either:

- **Reject H_0** (evidence for a difference/relationship), or
- **Fail to reject H_0** (not enough evidence, though this doesn't prove there's no difference).

Step Two:
Sample Size

When planning an audit or service evaluation, it's natural to ask: *"How many patients/samples do I need?"*

- Too few samples = less time/work/cost but the results may be unreliable or miss patterns (false negatives).
- Too many samples = lots of time/work/cost, and the results may detect tiny differences that are statistically significant but not clinically important.

The full answer can get very technical (statistical power, effect sizes, error rates), but for most community audits you just need a practical rule of thumb.

- **20–30** samples per group is generally the minimum for simple averages and to spot obvious issues.
- **50–100+** per group gives more robust comparisons between groups.
- For categorical outcomes, aim for at least **5–10** samples in each category.

Here's some example case studies:

- **In a cataract referral audit:** 5 cases could be distorted by one unusual patient; 50 cases will give a clearer pattern.
- **Optometrist record keeping audit:** 10 patients give a rough average; 40–50 patients make conclusions more trustworthy.
- **Patient satisfaction survey:** At least 30–40 responses are needed so one or two very un/happy patients don't dominate the results.

In community audits, don't get lost in power calculations. Collect enough data to see the trend (≈30+ patients per group), report your sample size, and keep the focus on whether findings are clinically meaningful.

Step Three:
Selecting the Right Statistical Test

So, you've collected your data. Which inferential test you use depends on your data type, whether it is parametric (roughly follows a normal distribution), the number of groups, and whether the data is paired.

Is the data parametric? (i.e. does the data follow a normal distribution?)

- **Parametric tests** are more powerful for numerical data, but assume the data is approximately normal (i.e. low skew).
- **Non-parametric tests:** More flexible, don't assume normality, and are suitable for skewed or ordinal data.

How many groups are you comparing?

- **Two groups:** This is usually the case in the typical scenario where you are comparing before/after an intervention.
- **Three or more groups:** For more complicated comparisons, you'll need to use different statistical tests.

Are your groups independent or related?

- **Independent groups:** Different sets of patients, e.g. comparing average satisfaction in a group using multifocal lens A vs a different group using multifocal lens B.
- **Related (paired) samples:** The same patients measured twice (or matched pairs), usually before and after an intervention, e.g. comparing IOP in the same patients before and after starting glaucoma treatment.

Step Four:
Perform your Statistical Test

While Microsoft Excel is not a dedicated statistical platform, it offers a range of built-in functions that allow users to perform basic inferential tests. For more advanced or large-scale analyses, it's often better to use specialised statistical software such as SPSS or JASP. Alternatively, the programming language R is widely used in academia and industry for statistical computing and visualisation.

However, for the purposes of this article, we'll focus on what's achievable directly within Excel. Below is a summary of common statistical tests categorized by data type, along with the relevant Excel functions or workarounds.

Data Analysis Toolpak: To access ANOVA and other statistical tests in Excel, enable the Toolpak via:

File → Options → Add-ins → Manage: Excel Add-ins → Go → check Analysis Toolpak

Non-parametric tests (like Mann-Whitney or Kruskal-Wallis) often require manual steps or external tools, as Excel doesn't natively support them. Fisher's Exact Test is typically used for small sample sizes in 2x2 contingency tables and is best handled in R or online calculators.

Nominal

- **Chi-squared:** CHISQ.TEST(actual_range, expected_range)
- **Fisher's Exact:** Not natively supported in Excel. Use online calculator or R.

Ordinal

- **Mann-Whitney U:** No direct formula. Use Rank + manual calculation or an external tool.
- **Kruskal-Wallis:** Not supported directly - requires ranking and manual steps, or the additional of an external tool

Discrete

- **Unpaired t-test:** T.TEST(array1, array2, tails, 2)
- **Paired t-test:** T.TEST(array1, array2, tails, 1)
- **ANOVA:** Use Data Analysis Toolpak → ANOVA: Single Factor

Continuous

- **T-test:** Same as above (T.TEST…)
- **ANOVA:** Use Data Analysis Toolpak → ANOVA: Single or Two-Factor
- **Mann-Whitney U:** See Ordinal; manual ranking required
- **Kruskal-Wallis:** See Ordinal; manual ranking required
- **Wilcoxon signed-rank:** Not supported directly. Manual ranking and difference signs required.

Step Five:
P-Values – What Do They Mean?

P-values represent the probability that the observed effect happened by random chance alone. After running your test, you'll get a p-value of between 0 and 1, or 0%-100%.

- If $p < 0.05$: The difference is in most cases considered "statistically significant", as there is a 95% probability that the result did not occur by chance.
- If $p \geq 0.05$: This is usually considered not statistically significant, meaning that there is not enough evidence to reject H_0.

However, a word of caution, a result can be statistically significant but not clinically meaningful. For example, a 0.02 LogMAR improvement in VA may be significant with a large dataset, but it wouldn't necessarily change your patient care decisions.

Further Resources

This article aims to give you an understanding of how to go about conducting a robust audit or service evaluation but cannot hope to give you the full set of statistical tools for every eventuality. The best way to learn more is by doing and by exploring real-world examples.

YouTube Channels:

- StatQuest with Josh Starmer: Excellent, engaging explanations of statistical concepts. Search for specific tests (t-tests, chi-squared).
- Khan Academy (Statistics): Comprehensive and well-structured lessons on basic to intermediate statistics.

Online Tutorials & Textbooks:

- Many university statistics departments offer free online materials. Search for "introductory statistics for health sciences."
- OpenLearn: Free course on medical statistics.

Software:

- Microsoft Excel: Capable of basic descriptive statistics and charting. Has an "Analysis ToolPak" for some inferential tests (though caution is needed for complex analyses).
- R or Python: Free, powerful programming languages for statistics, but with a steeper learning curve.
- Jamovi / JASP: Free, open-source statistical software with a user-friendly interface, similar to commercial packages like SPSS. Excellent for learning and conducting analyses.

Books:

- Introduction to Medical Statistics, by M.Bland
- Medical Statistics Made Easy, by M.Harris and G.Taylor

Finally, don't be intimidated by statistics. Start with simple questions, understand your data, use the right tools, and always interpret your findings in the context of patient care and service improvement.

For further advice or feedback on the content of this document, please contact info@locsu.co.uk

2 Woodbridge Street,
London EC1R 0DG
e info@locsu.co.uk
t 020 7549 2051
locsu.co.uk

Version 1: 12/25

LOCSU