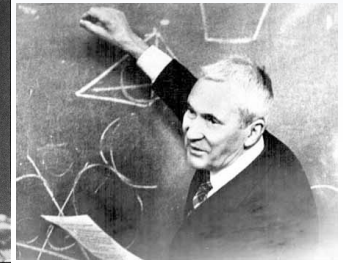
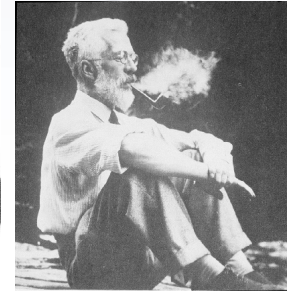
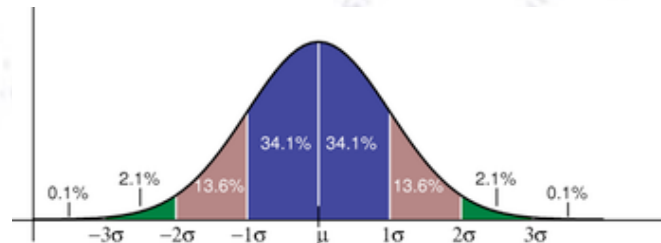


Applied Statistics

On Trial Factors

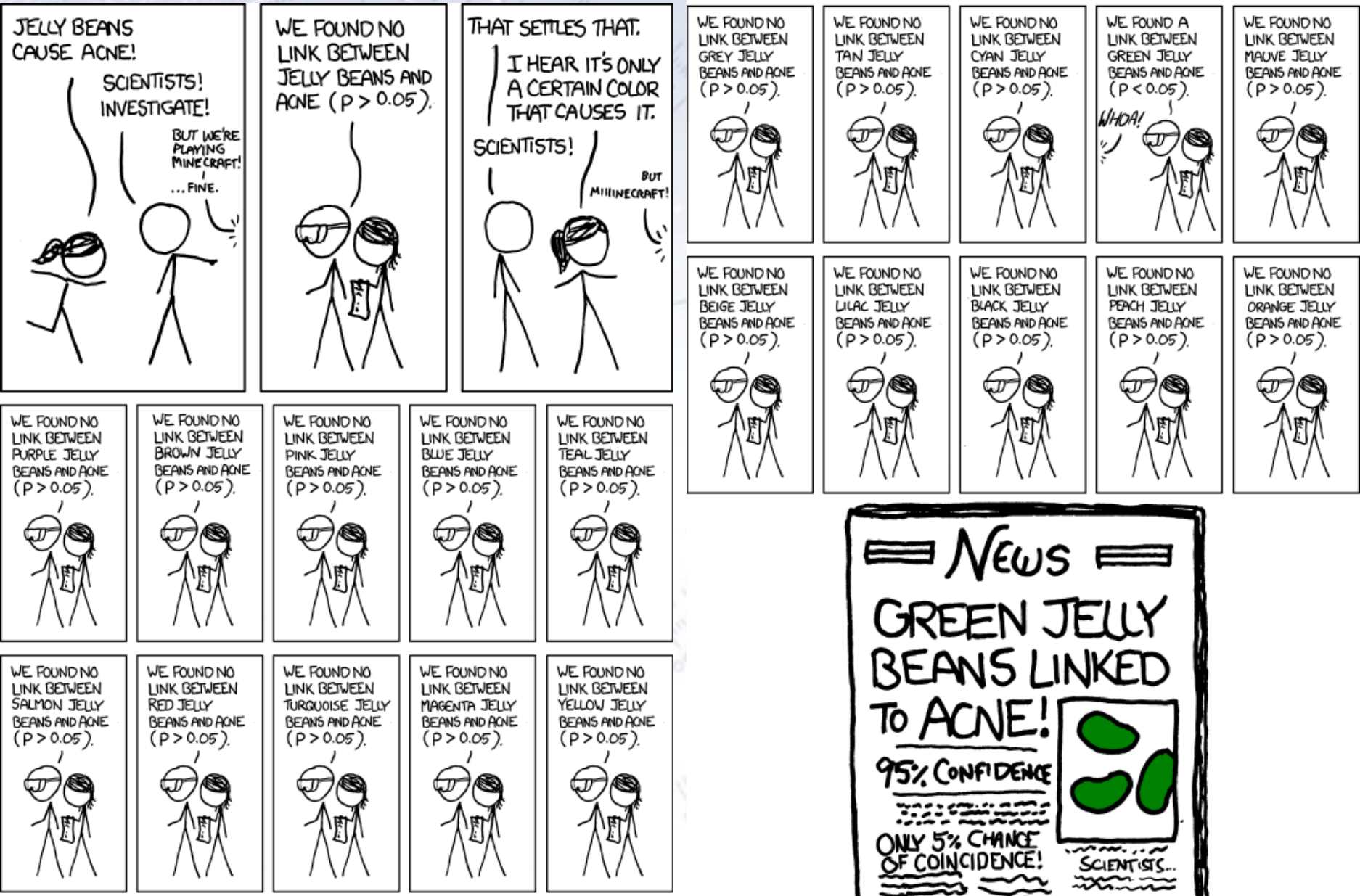


Troels C. Petersen (NBI)

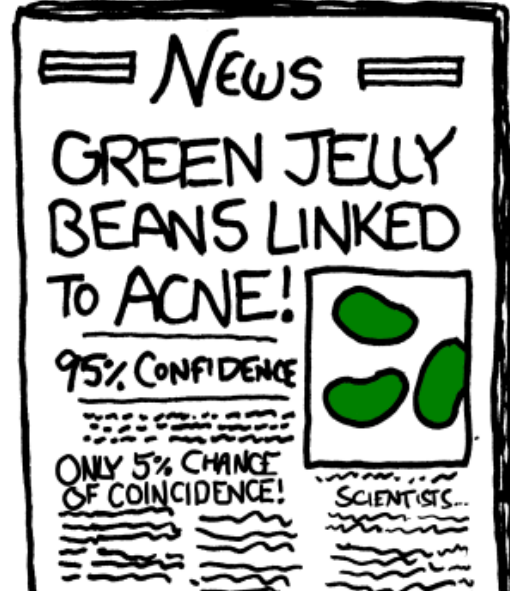
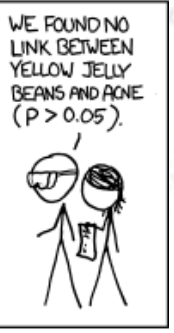
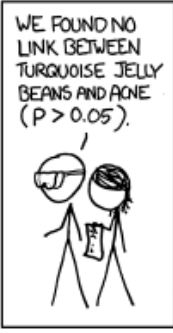
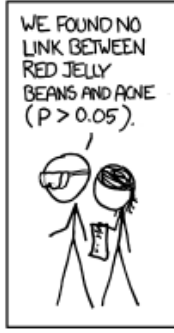
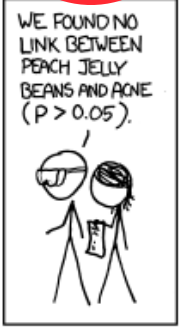
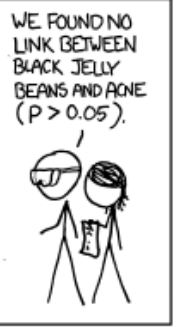
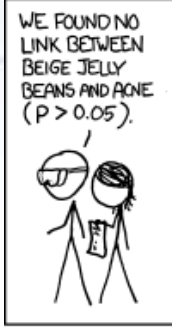
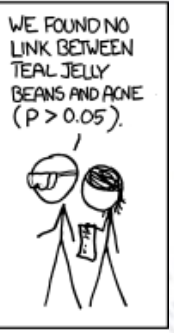
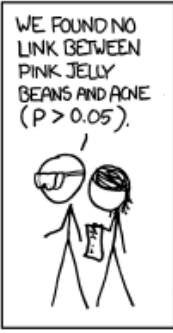
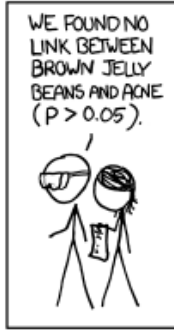
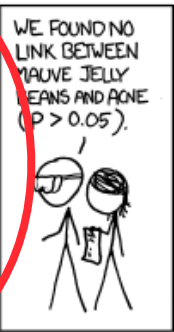
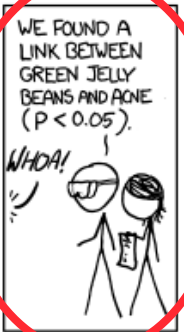
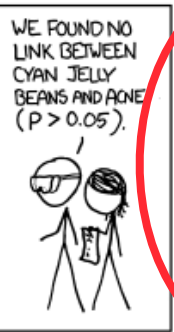
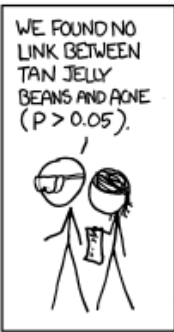
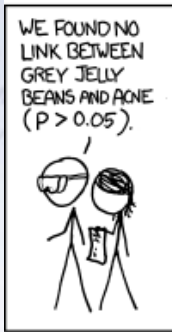


"Statistics is merely a quantisation of common sense"

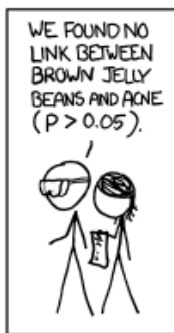
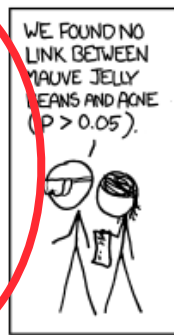
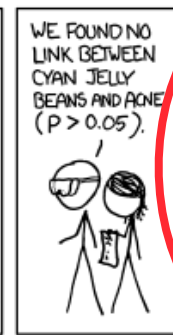
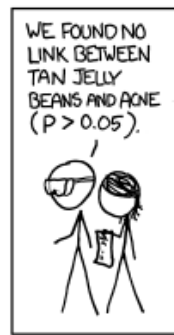
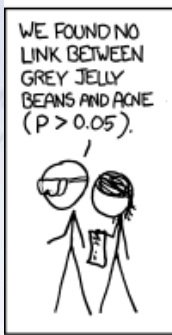
XKCD on statistics



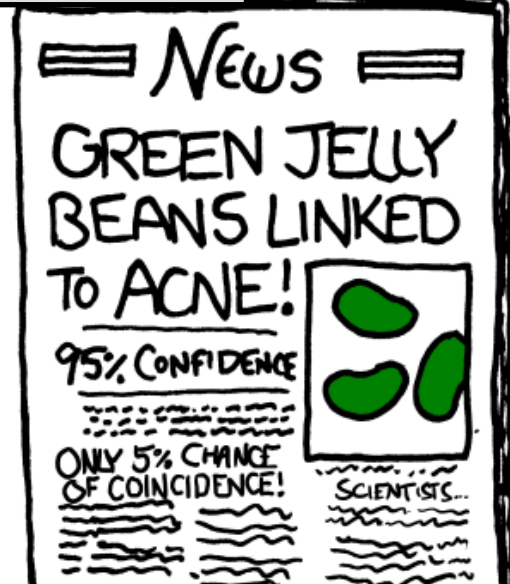
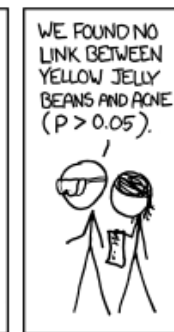
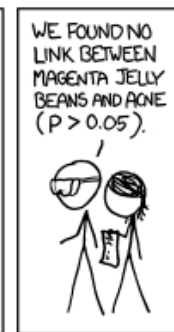
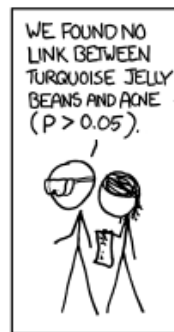
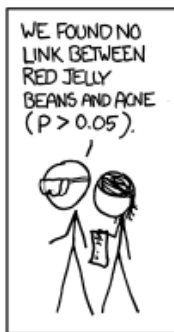
XKCD on statistics



XKCD on statistics



Trial factor: 20



Trial factor / Look-Elsewhere Effect

“If you look enough times or places, you will find something unlikely”

The “Look-Elsewhere Effect” refers to observing an **apparent** statistically significant observation, which has arisen from searching a large parameter space (i.e. many places).

To account for this, one uses a **trial factor**, which is the ratio between the probability of observing a possible excess at some fixed point, to the probability of observing it anywhere in the range.

The significance of the (fitted) amplitude tells you the **local significance**. As you might be searching in many places, this reduces your certainty to the **global significance**:

$$p_{\text{global}} = 1 - (1 - p_{\text{local}})^N \approx N \text{ local}$$

Thus, the global significance is (roughly) reduced by the trial factor.

A good paper with discussion of statistical treatment: <https://arxiv.org/abs/1005.1891>

Example case

A Swedish study in 1992 tried to determine whether or not power lines caused some kind of poor health effects. The researchers surveyed everyone living within 300 m of high-voltage power lines over a 25-year period and looked for statistically significant increases in rates of **over 800 ailments**.

The study found that the incidence of childhood leukemia was four times higher among those that lived closest to the power lines, and it spurred calls to action by the Swedish government.

Example case

A Swedish study in 1992 tried to determine whether or not power lines caused some kind of poor health effects. The researchers surveyed everyone living within 300 m of high-voltage power lines over a 25-year period and looked for statistically significant increases in rates of **over 800 ailments**.

The study found that the incidence of childhood leukemia was four times higher among those that lived closest to the power lines, and it spurred calls to action by the Swedish government.

The problem with the conclusion, however, was that they failed to compensate for the **look-elsewhere effect**; in any collection of 800 random samples, it is likely that at least one will be at least 3 standard deviations above the expected value, **by chance alone**. Subsequent studies failed to show any links between power lines and childhood leukemia, neither in causation nor even in correlation.

[Jon Palfreman, "Currents of fear" (1995-06-13), Frontline, PBS,]