The t-distributions

In our previous examples we assumed that the standard deviation of a population was known but in most cases neither μ or σ is known.

If we don't know σ it seems natural to use sample standard deviation s, but this will result in less accurate standard error $\sigma_{\overline{x}}$ since we would use $\sigma_{\overline{x}} = s/\sqrt{n}$.

As a result, the z-statistic will be replaced with a statistic that accounts for the use of an estimated standard error. These new (larger) values are called the **t-values (or the Student's t-statistic).**

However, if the sample size n is sufficiently large, the difference between z-values and t-values becomes negligible.

We find t-values in tables like we did with z-values.

The t-table

Since the t-values are dependent on our sample size, critical values of t are arranged according to **degrees of freedom** df.

$$df = n - 1$$

For example, let's find the critical value of t for $\alpha = 0.025$ if $n = 28$.

Summary

The last line of the table is critical values of z.

The convention used today is to use the t-table for sample sizes up to n=100 and use the z-table (last line) for larger samples.

We do not use t-tables for proportions.

Use z-table when σ is known

Use t-table when σ is unknown

•

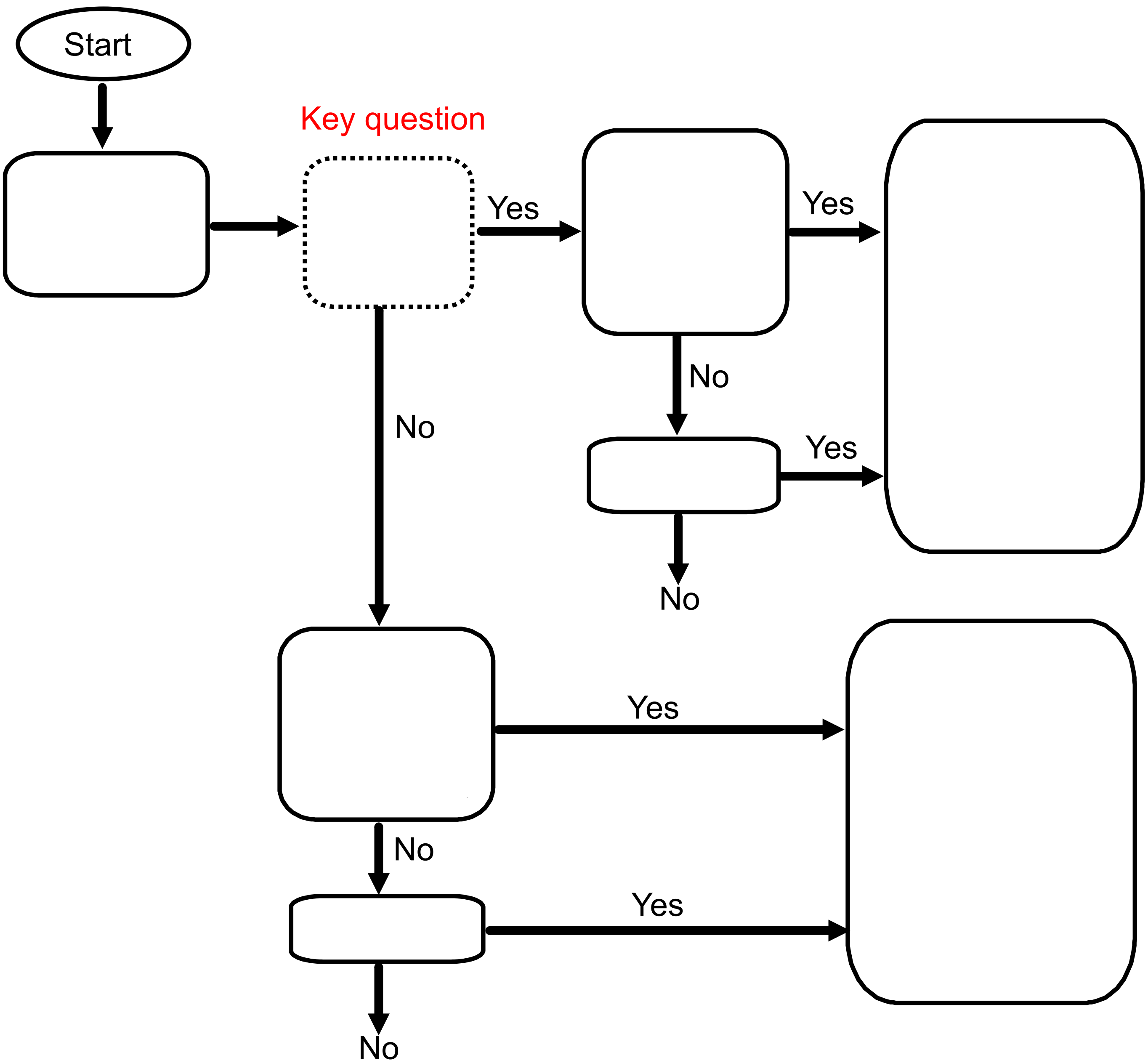
•

•

(Note: The text uses the notation $t(df, \alpha)$ instead of $t(\alpha)$)

Notice that it is difficult to control the error using sample size since we don't have a standard deviation until we actually have a sample.

z-statistic or t-statistic?



Example: Steel bars were produced with the following tensile strengths

305	287	306	331	339	353	255	304	323
337	282	292	267	263	285	290	257	332
295	299	303	300	311	300	321	307	357
305	281	359	331	369	289	351	269	311

Construct a 95% C.I.E. for mean tensile strength of steel bars (round to one decimal place).

Test of Hypothesis, σ unknown (t-test)

The steps for testing a hypothesis if σ is unknown are the same when we know σ except we use s instead of σ, the test statistic

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

and $t(\alpha)$ values with $df = n - 1$.

Example: Past data shows that monthly phone bill costs are normally distributed with mean $17.50. After an ad campaign a random sample of 21 households was taken finding $\overline{x} = \$19.10$ with $s = \$3.74$. Was the ad campaign effective? Test at 5% significance.