One-Sample Tests

Single Samples

Questions we can ask:

What is the mean value?

Is the mean value significantly different than expectation or theory?

What is the level of uncertainty associated with the mean value?

Assumptions

normally distributed

random sample from the population

no serial correlation

If the assumptions of normality are violated, the inferences of a parametric test (like Student's t test) will be violated.

In these cases, it is better to use a non-parametric technique (like Wilcoxon's signed-rank test).

A parametric test makes assumptions about the parameters of the distribution of the population from which the data are drawn.

Usually that a population is normally distributed, etc.

A non-parametric test makes fewer assumptions about the distribution.

But they often have less power than parametric tests.

If there is serial correlation in the data, then it is better to use a time series or mixed-effects model.

We won’t worry about this today but keep this in mind when working on data collected over time.

Student's t-test (parametric)

H_o: The true mean equals μ_o

H_A: The true mean does not equal μ_o

The one-sample t-test compares the mean of a random sample from a normal population with the population mean proposed in a null hypothesis.

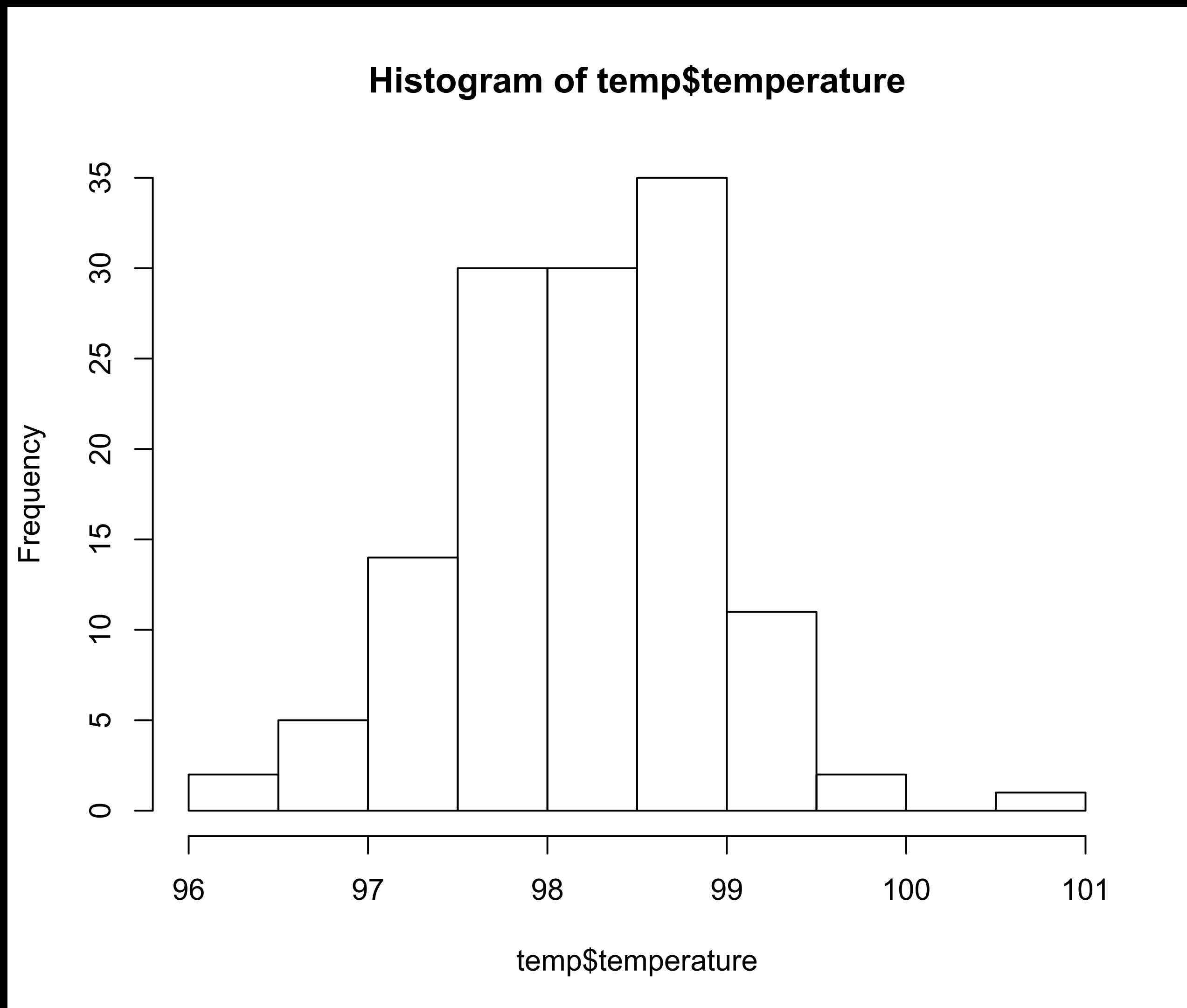
Body Temperature Data

What is the normal human body temperature? Is it 98.6 F, as we were taught in elementary school?

H_o: The true mean equals 98.6

H_A: The true mean does not equal 98.6

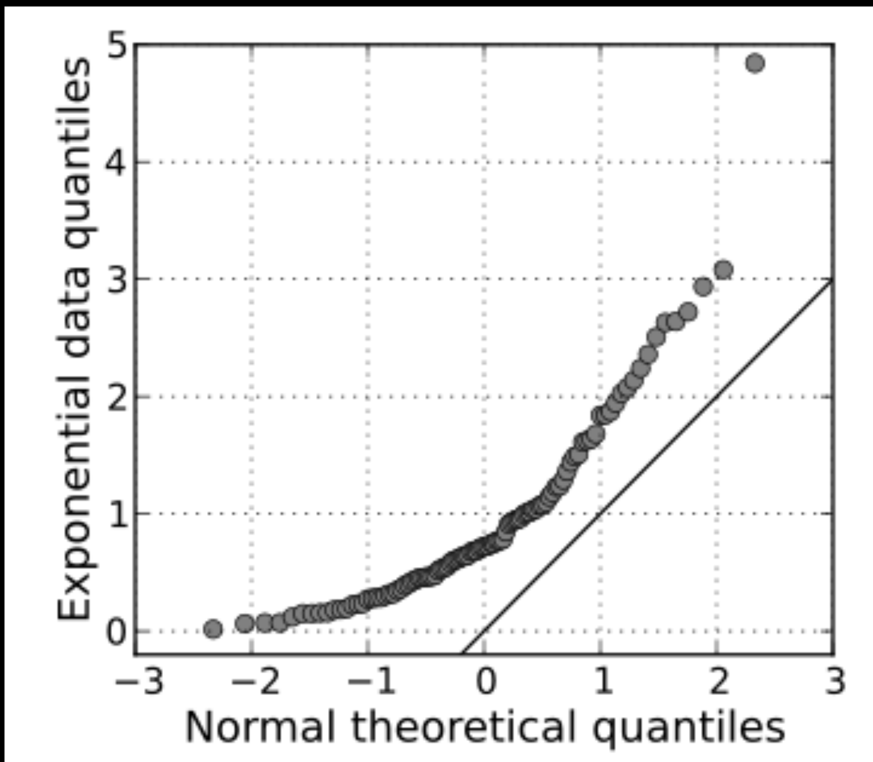
histogram of data



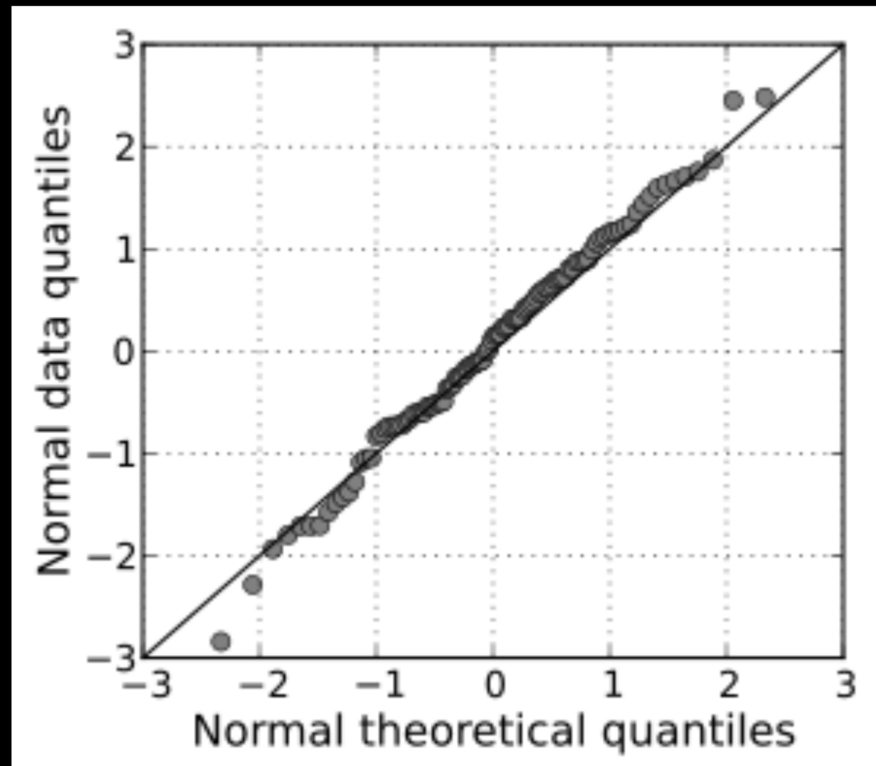
tests for normality

quantile-quantile plot: compares probability distributions by plotting their quantiles against another

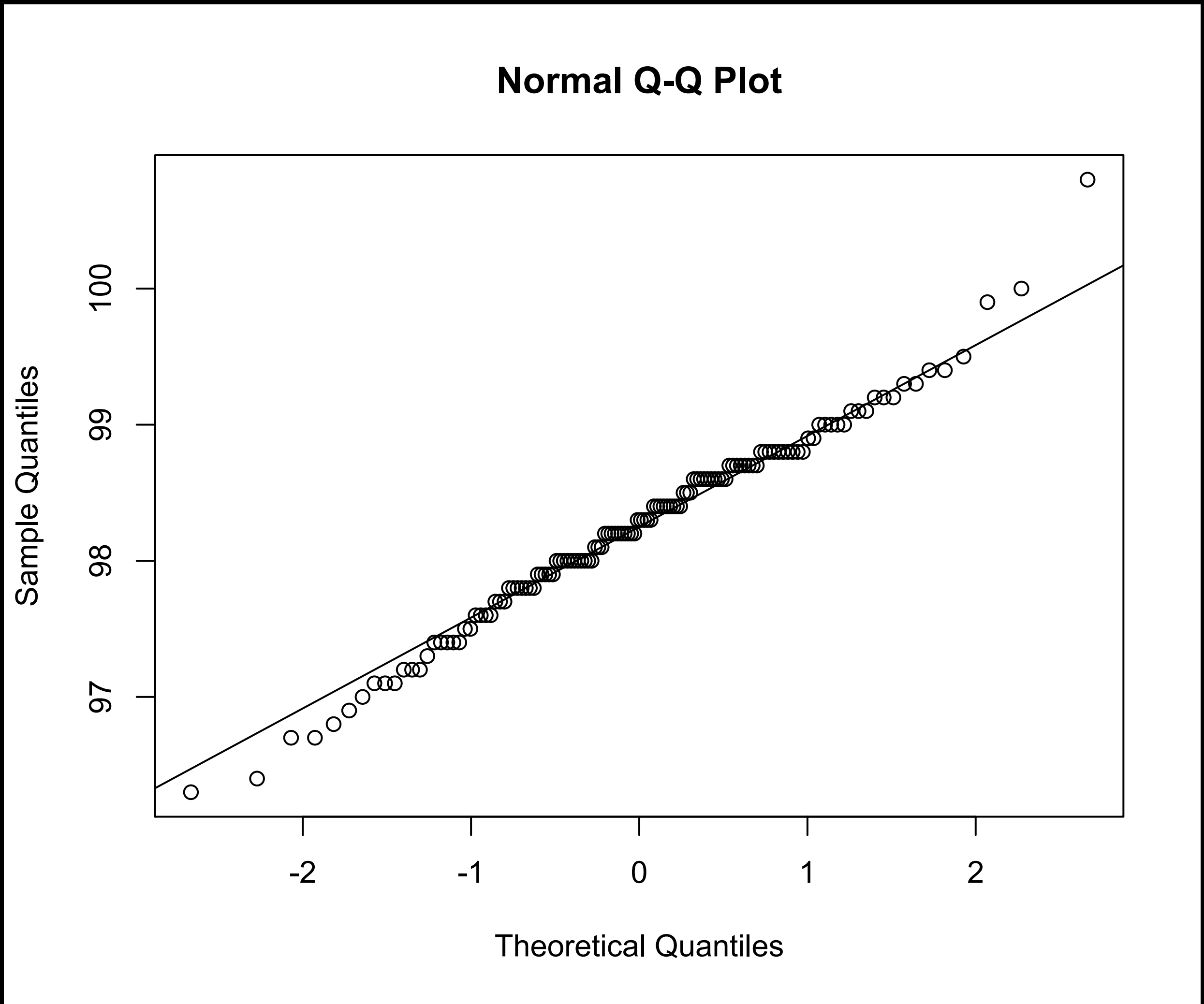
so, we can plot the quantiles of our data against the quantiles of a normal distribution—if the line is relatively straight, it is normal not normal



normal



test for normality



tests for normality

Shapiro-Wilk test: tests against the null hypothesis that the population is normal

if the result is significant, then we determine that our data are violate the assumptions of normality

Shapiro-Wilk Test

```
W = 0.9866, p-value = 0.2332
```

W is the test statistic.

If the p-value is less than the chosen level, then we reject the null hypothesis that our data are normally distributed.

there are other tests for normality

Anderson-Darling test

Martinez-Iglewicz test

Kolmogorov-Smirnov test

D'Agostino Omnibus test

t distribution

used when population standard deviation unknown

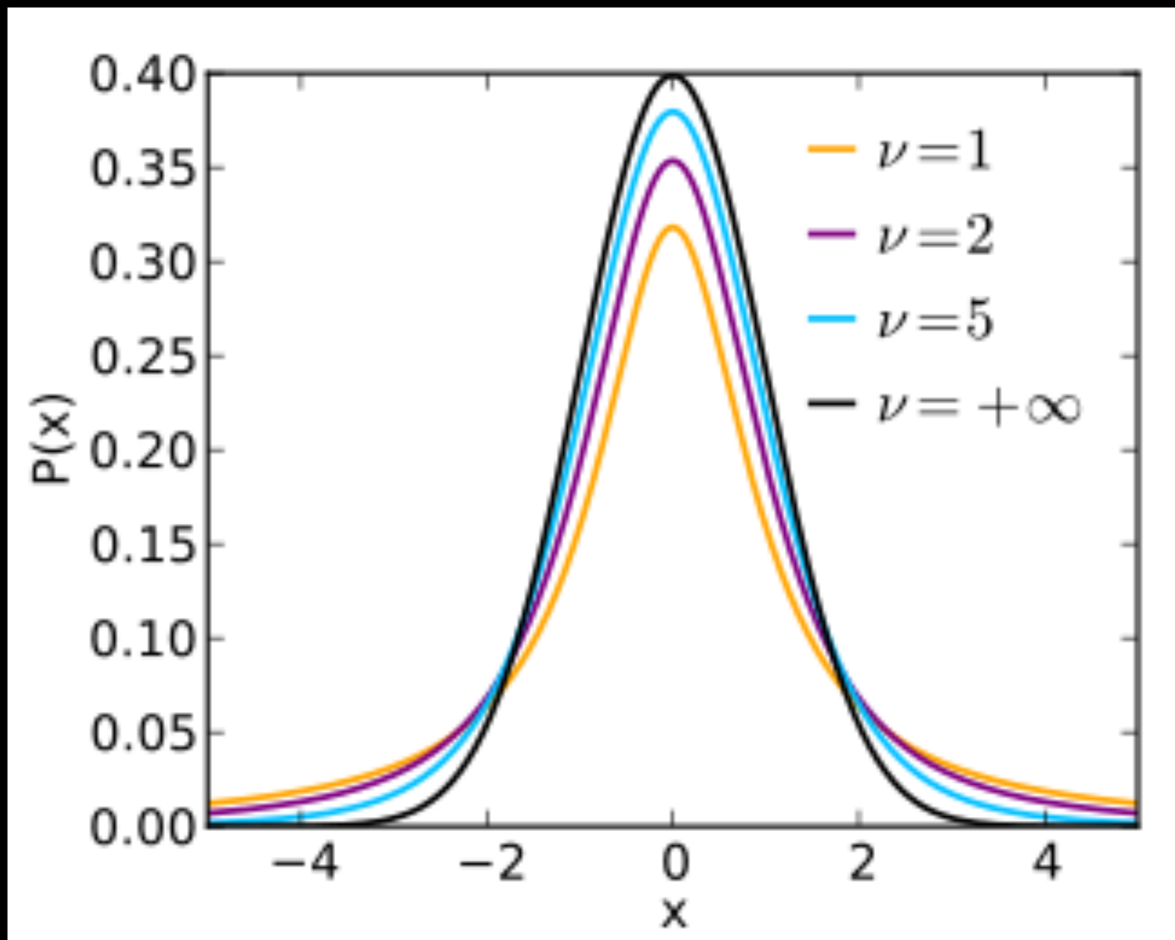
similar to standard normal, with fatter tails

as sample size increases, becomes more like standard normal

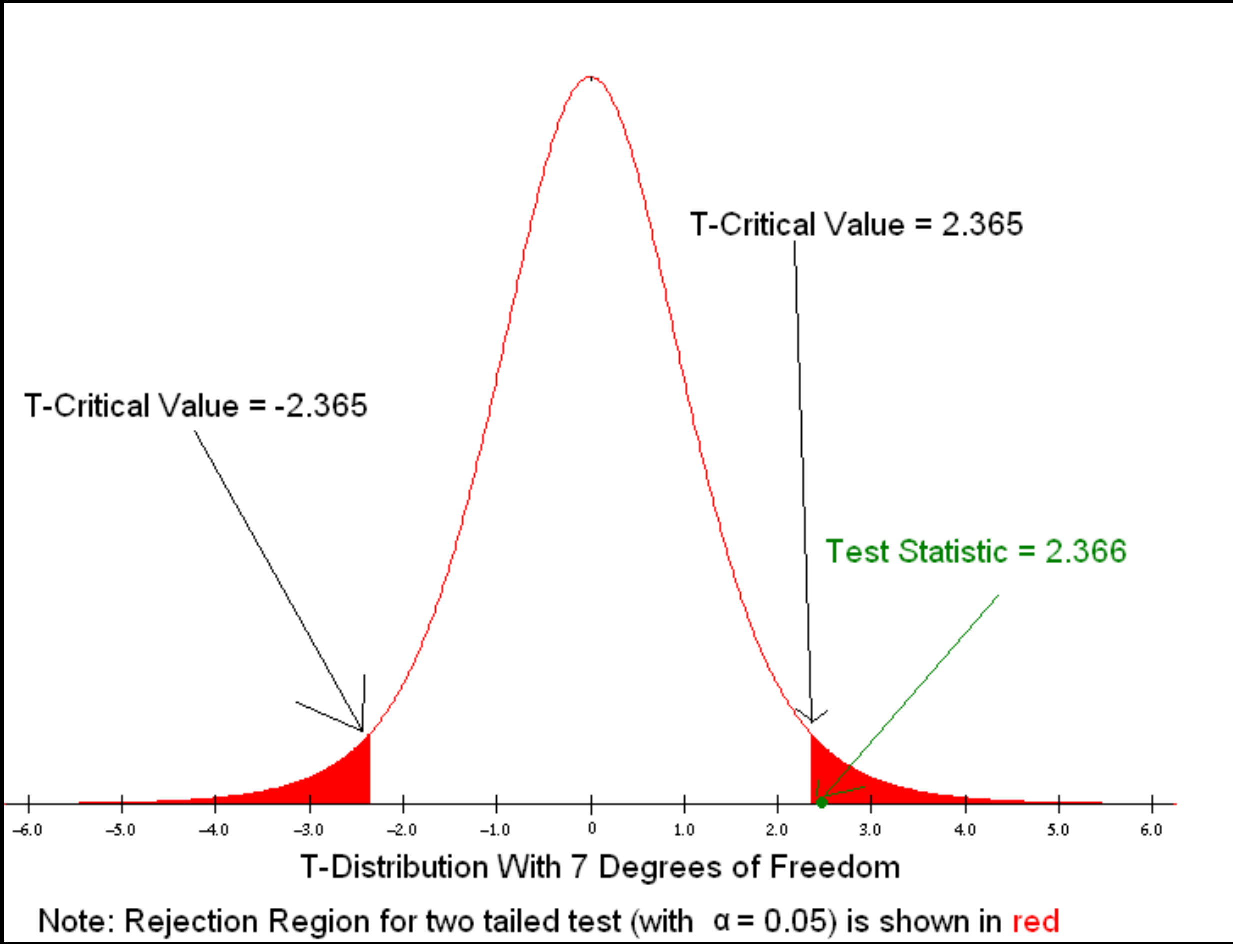
has degrees of freedom, determined by the sample size

$df = n - 1$

t distribution



critical values of the t distribution



Body Temperature

t = -5.4548
df = 129
p-value = 2.411e-07
confidence interval: 98.12200 98.37646
sample estimates: mean of x 98.24923

So, we reject the null hypothesis that the population mean of human body temperature is equal to 98.7F.

Confidence Intervals

range of values surrounding the sample estimate that is likely to contain the population parameter

quantifies uncertainty about the value of the parameter

what if the data aren't normal?

ignore violations of the assumptions

try a transformation, then test for normality

turn to nonparametric tests

nonparametric tests

usually based on ranks of the data points rather than the actual values of the data

if the null hypothesis were true, these ranks would, on average, be about the same

not affected as much by outliers

sign test

assesses whether the median of a population equals a null hypothesized value

very little power compared to one-sample *t*-test because it discards most of the information in the data

Wilcoxon signed-rank test

like sign test except retains information about the magnitude of the value above the median

but, requires the distribution to be symmetric (no skew)

Two-Sample Tests

two-sample tests assumptions

follow normal probability distribution

variances of two populations are equal

two samples are independent

random samples

drug vs. placebo

What is our null hypothesis?

Two Samples

compare variances (Fisher's F test, `var.test`)

compare two sample means with normal errors (Student's *t*-test, `t.test`)

compare two means with non-normal errors (Wilcoxon's rank test, `wilcox.test`)

test for normality

compare two variances

Fisher's *F* test: divide larger variance by smaller, look to see if ratio is significantly different than 1

assumes normality—if not normal, use Levene's test

for multiple samples

Fligner-Killen test

```
fligner.test()
```

Bartlett test

```
bartlett.test()
```

interpreting var.test

```
F test to compare two variances
data:  drugs$drug and drugs$placebo
F = 1.9791, num df = 9, denom df = 9,
p-value = 0.3237
alternative hypothesis: true ratio of
variances is not equal to 1
95 percent confidence interval: 0.491579
  7.967821
sample estimates:ratio of variances
  1.979094
```

Welch's two-sample t-test

so, we have equality of variances, so we don't need to use Welch's two-sample test (which accounts for nonequality of variances)

adjusts the degrees of freedom

Welch is the default in R, so we have to turn this off (`var.equal=TRUE`)

interpreting t.test

```
Two Sample t-test
data:  drugs$drug and drugs$placebo
t = -0.5331, df = 18,
p-value = 0.6005
alternative hypothesis: true difference
  in means is not equal to 0
95 percent confidence interval:
  -4.446765  2.646765
sample estimates:mean of x mean of y
  11.4      12.3
```

So, we accept the null hypothesis that the drug and the placebo have the same mean recovery time.

Wilcoxon rank-sum test

used in place of two-sample *t*-test when the normal distribution assumption can not be met

compares medians of two groups

distributions of the two groups must have same shape (variance and skew)

Mann-Whitney *U*-test has a differently calculated test statistic but will give you equivalent results

paired tests

when there is an association between data points in each group

violates assumption of independence

ex: before and after, husbands and wives, matched treatments and controls, upstream and downstream

paired tests

sample size must be equal

even though there are two samples, you work with one sample composed of standard differences