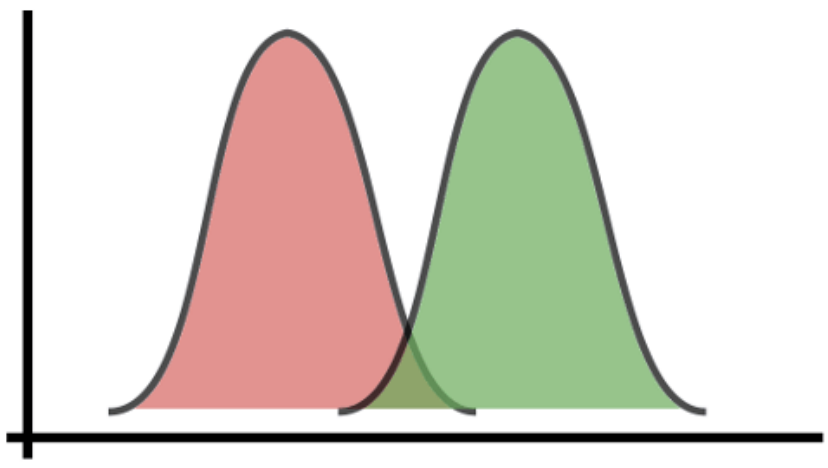

julia

Statistics with Julia

- Its all very well generating myriad statistics characterising your data.
- How do you know whether or not those statistics are telling you something interesting? Hypothesis Tests.
- To that end, well be looking at the HypothesisTests package today.

Statistics with Julia

```
Julia> using HypothesisTests
julia> using Distributions

julia> srand(357)
julia> x1 = rand(Normal(), 1000);
julia> x2 = rand(Normal(0.5, 1), 1000);

julia> # 25% success rate on samples of size 100
julia> x3 = rand(Binomial(100, 0.25), 1000);

julia> # 50% success rate on samples of size 50
julia> x4 = rand(Binomial(50, 0.50), 1000);
julia> x5 = rand(Bernoulli(0.25), 100) .== 1;
```

Statistics with Julia

- Well apply a one sample t-test to x1 and x2. The output below indicates that x2 has a mean which differs significantly from zero while x1 does not.
- This is consistent with our expectations based on the way that these data were generated.
- Im impressed by the level of detail in the output from `OneSampleTTest()`: different aspects of the test are neatly broken down into sections (population, test summary and details) and there is automated high level interpretation of the test results.

Statistics with Julia

```
julia> t1 = OneSampleTTest(x1)
One sample t-test
-----------------
Population details:
parameter of interest:   Mean
value under h_0:         0
point estimate:          -0.013027816861268473
95% confidence interval: (-0.07587776077157478,0.049822

Test summary:
outcome with 95% confidence: fail to reject h_0
two-sided p-value:           0.684269269639374 (not s
```

Statistics with Julia

```
Details:
number of observations:   1000
t-statistic:              -0.40676289562651996
degrees of freedom:       999
empirical standard error: 0.03202803648352013
julia> t2 = OneSampleTTest(x2)
One sample t-test
```

Statistics with Julia

```
Population details:
parameter of interest:   Mean
value under h_0:         0
point estimate:          0.507852246706
95% confidence interval: (0.4468203,0.568884)

Test summary:
outcome with 95% confidence: reject h_0
two-sided p-value:           2.62561601163e-53
Details:
number of observations:   1000
t-statistic:              16.32883382693
degrees of freedom:       999
empirical standard error: 0.03110156255427
```

Statistics with Julia

- Using `pvalue()` we can further interrogate the p-values generated by these tests.
- The values reported in the output above are for the two-sided test, but we can look specifically at values associated with either the left- or right tails of the distribution.
- This makes the outcome of the test a lot more specific.

Statistics with Julia

```
julia> pvalue(t1)
0.6842692696393744

julia> pvalue(t2)
2.6256160116367554e-53

julia> pvalue(t2, tail = :left)                    #
1.0

julia> pvalue(t2, tail = :right)
1.3128080058183777e-53
```

Statistics with Julia

The associated confidence intervals are also readily accessible. We can choose between two-sided or left/right one-sided intervals as well as change the significance level.

```
julia> ci(t2, tail = :both)
(0.4468203610006495 4,0.5688841324132342)

julia> ci(t2, tail = :left)
(-Inf,0.5590572480083876)

julia> ci(t2, 0.01, tail = :right)
(0.4353829181883160 4,Inf)
```

Statistics with Julia

- As a second (and final) example well look at `BinomialTest()`.
- There are various ways to call this function. First, without looking at any particular data, well check whether 25 successes from 100 samples is inconsistent with a 25% success rate (obviously not and, as a result, we fail to reject this hypothesis).

Statistics with Julia

```
julia> BinomialTest(25, 100, 0.25)
Binomial test
-------------
Population details:
parameter of interest:   Probability of success
value under h_0:         0.25
point estimate:          0.25
95% confidence interval: (0.16877973809,0.3465524957)

Test summary:
outcome with 95% confidence: fail to reject h_0
two-sided p-value:           1.0 (not signficant)

Details:
number of observations: 100
number of successes:    25
```

Statistics with Julia

Next well see whether the Bernoulli samples in x5 provide contradictory evidence to an assumed 50% success rate (based on the way that x5 was generated we are not surprised to find an infinitesimal p-value and the hypothesis is soundly rejected).

```
julia> BinomialTest(x5, 0.5)
Binomial test
-------------
Population details:
parameter of interest:   Probability of success
value under h_0:         0.5
point estimate:          0.18
95% confidence interval: (0.1103112291532605,0.2694770859

Test summary:
outcome with 95% confidence: reject h_0
two-sided p-value:           6.147806615048005e-11 (extreme

Details:
number of observations: 100
number of successes:    18
```

Statistics with Julia

- There are a number of other tests available in this package, including a range of non-parametric tests which I have not even mentioned above.
- Certainly HypothesisTests should cover most of the bases for statistical inference. For more information, read the extensive documentation.