t-Test Statistics

Overview of Statistical Tests
Assumption: Testing for Normality
The Student's *t*-distribution
Inference about one mean (one sample *t*-test)
Inference about two means (two sample *t*-test)
Assumption: *F*-test for Variance
Student's *t*-test
- For homogeneous variances
- For heterogeneous variances

Statistical Power

Overview of Statistical Tests

During the design of your experiment you must specify what statistical procedures you will use.

You require at least 3 pieces of info:

Type of Variable
Number of Variables
Number of Samples

Then refer to end-papers of Sokal and Rohlf (1995)
-REVIEW-

Assumptions

Virtually every statistic, parametric or nonparametric, has *assumptions* which must be met prior to the application of the experimental design and subsequent statistical analysis.

We will discuss specific assumptions associated with individual tests as they come up.

Virtually all parametric statistics have an assumption that the data come from a population that follows a known distribution.
Most of the tests we will evaluate in this module require a normal distribution.

Assumption: Testing for Normality

Review and Commentary:

D'Agostino, R.B., A. Belanger, and R.B. D'Agostino. 1990. A suggestion for using powerful and informative tests of normality. The American Statistician 44: 316-321.

(See Course Web Page for PDF version.)

Most major normality tests have corresponding R code available in either the base stats package or affiliated package. We will review the options as we proceed.

Normality

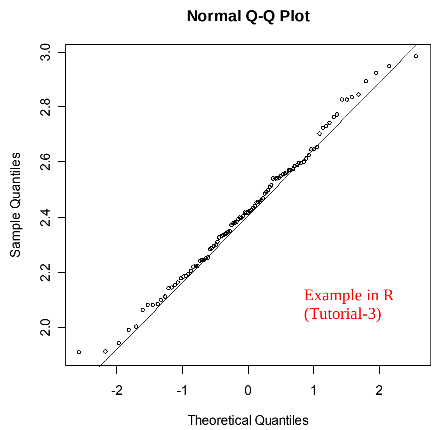
There are 5 major tests used:

Shapiro-Wilk W test
Anderson-Darling test
Martinez-Iglewicz test
Kolmogorov-Smirnov test
D'Agostino Omnibus test

NB: Power of all is weak if $N < 10$

Shapiro-Wilk *W* test

Developed by Shapiro and Wilk (1965).
One of the most powerful overall tests.
It is the ratio of two estimates of variance (actual calculation is cumbersome; adversely affected by ties).
Test statistic is *W*; roughly a measure of the straightness of the quantile-quantile plot.
The closer W is to 1, the more normal the sample is.
Available in R and most other major stats applications.



Anderson-Darling test

Developed by Anderson and Darling (1954).
Very popular test.
Based on EDF (empirical distribution function) percentile statistics.
Almost as powerful as Shapiro-Wilk *W* test.

Martinez-Iglewicz Test

Based on the median & robust estimator of dispersion.
Very powerful test.
Works well with small sample sizes.
Particularly useful for symmetrically skewed samples.
A value close to 1.0 indicates normality.
Strongly recommended during EDA.
Not available in R.

Kolmogorov-Smirnov Test

Calculates expected normal distribution and compares it with the observed distribution.
Uses cumulative distribution functions.
Based on the max difference between two distributions.
Poor for discrimination below *N* = 30.
Power to detect differences is low.
Historically popular.
Available in R and most other stats applications.

10

D'Agostino et al. Tests

Based on coefficients of skewness ($\sqrt{b_1}$) and kurtosis (b_2).

If normal, $\sqrt{b_1}=1$ and $b_2=3$ (tests based on this).

Provides separate tests for skew and kurt:

- Skewness test requires $N \geq 8$
- Kurtosis test best if $N > 20$

Provides combined Omnibus test of normality.

Available in R.

11

The *t*-Distribution

t-distribution is similar to *Z*-distribution

Note similarity:

$$Z = \frac{\bar{y} - \mu}{\sigma / \sqrt{N}} \quad vs. \quad t = \frac{\bar{y} - \mu}{S / \sqrt{N}}$$

The functional difference is between σ and S.
Virtually identical when *N* > 30.

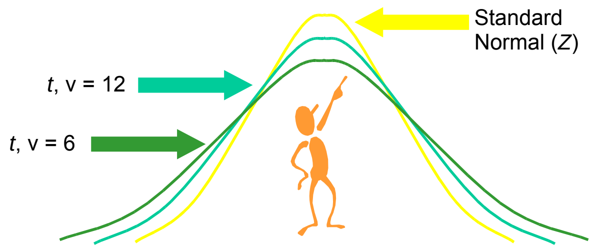
Much like *Z*, the *t*-distribution can be used for inferences about μ.

One would use the *t*-statistic when σ is not known and S is (the general case).

12

The *t*-Distribution

See Appendix
Statistical Table C



13

One Sample *t*-test

- Assumptions -

The data must be continuous.

The data must follow the normal probability distribution.

The sample is a simple random sample from its population.

14

One Sample *t*-test

$$t = \frac{\bar{y} - \mu}{S / \sqrt{N}}$$

$$\bar{y} - t_{\alpha(2), df} SE_{\bar{y}} \leqslant \mu \leqslant \bar{y} + t_{\alpha(2), df} SE_{\bar{y}}$$

$$\frac{df\ s^2}{\chi^2_{\alpha/2, df}} \leqslant \sigma^2 \leqslant \frac{df\ s^2}{\chi^2_{1-\alpha/2, df}}$$

15

One Sample *t*-test

- Example -

Twelve (N = 12) rats are weighed before and after being subjected to a regimen of forced exercise.

Each weight change (g) is the weight after exercise minus the weight before:

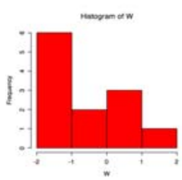
1.7, 0.7, -0.4, -1.8, 0.2, 0.9, -1.2, -0.9, -1.8, -1.4, -1.8, -2.0

H_0: $\mu = 0$
H_A: $\mu \neq 0$

One Sample *t*-test

- Example -

```
> W<-c(1.7, 0.7, -0.4, -1.8, 0.2, 0.9, -1.2,
-0.9, -1.8, -1.4, -1.8, -2.0)

> summary(W)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -2.000  -1.800  -1.050  -0.650   0.325   1.700

> hist(W, col="red")

> shapiro.test(W)

      Shapiro-Wilk normality test

data:  W
W = 0.8949, p-value = 0.1364
```



One-sample *t*-test using R

```
> W<-c(1.7, 0.7, -0.4, -1.8, 0.2, 0.9, -1.2, -0.9,
+ -1.8, -1.4, -1.8, -2.0)
> W
 [1]  1.7  0.7 -0.4 -1.8  0.2  0.9 -1.2 -0.9 -1.8 -1.4
-1.8 -2.0

> t.test(W, mu=0)

        One Sample t-test

data:  W
t = -1.7981, df = 11, p-value = 0.09964
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.4456548  0.1456548
sample estimates:
mean of x
    -0.65
```

One Sample *t*-test

For most statistical procedures, one will want to do a post-hoc test (particularly in the case of failing to reject H_0) of the *required sample size* necessary to test the hypothesis.

For example, how large of a sample size would be needed to reject the null hypothesis of the one-sample *t*-test we just did?

Sample size questions and related error rates are best explored through a **power analysis**.


```
> power.t.test(n=15, delta=1.0, sd=1.2523, sig.level=0.05,
type="one.sample")

     One-sample t test power calculation

              n = 15
          delta = 1
             sd = 1.2523
      sig.level = 0.05
          power = 0.8199317
    alternative = two.sided

> power.t.test(n=20, delta=1.0, sd=1.2523, sig.level=0.05,
type="one.sample")

     One-sample t test power calculation

              n = 20
          delta = 1
             sd = 1.2523
      sig.level = 0.05
          power = 0.9230059
    alternative = two.sided

> power.t.test(n=25, delta=1.0, sd=1.2523, sig.level=0.05,
type="one.sample")

     One-sample t test power calculation

              n = 25
          delta = 1
             sd = 1.2523
      sig.level = 0.05
          power = 0.9691447
```


Two Sample *t*-test

- Assumptions -

The data are continuous (not discrete).

The data follow the normal probability distribution.

The variances of the two populations are equal. (If not, the Aspin-Welch Unequal-Variance test is used.)

The two samples are independent. There is no relationship between the individuals in one sample as compared to the other.

Both samples are simple random samples from their respective populations.

Two Sample *t*-test

Determination of which two-sample *t*-test to use is dependent upon first testing the variance assumption:

Two Sample *t*-test for Homogeneous Variances

Two-Sample *t*-test for Heterogeneous Variances

Variance Ratio *F*-test

- Variance Assumption -

Must explicitly test for homogeneity of variance

H_o: $S_1^2 = S_2^2$
H_a: $S_1^2 \neq S_2^2$

Requires the use of *F*-test which relies on the *F*-distribution.

$F_{calc} = S^2_{max} / S^2_{min}$

Get F_{table} at N-1 df for each sample
If $F_{calc} < F_{table}$ then fail to reject H_o.

Variance Ratio F-test

- Example -

Suppose you had the following sample data:

Sample-A	$S_a^2 = 16.27$	*N* = 12
Sample-B	$S_b^2 = 13.98$	*N* = 8

$F_{calc} = 16.27/13.98 = 1.16$
$F_{table} = 3.603$ (df = 11,7)

Decision: $F_{calc} < F_{table}$ therefore fail to reject Ho.

Conclusion: the variances are homogeneous.

Variance Ratio F-test
- WARNING -

★ Be careful! ★

The *F*-test for variance requires that the two samples are drawn from normal populations (i.e., must test normality assumption first).

If the two samples are not normally distributed, do not use Variance Ratio *F*-test !

Use the Modified Levene Equal-Variance Test.

Modified Levene Equal-Variance Test

$$z_{1j} = |x_j - Med_x|$$

$$z_{2j} = |y_j - Med_y|$$

First, redefine all of the variates as a function of the difference with their respective median.

Then perform a two-sample ANOVA to get *F* for redefined values.

Stronger test of homogeneity of variance assumption. Not currently available in R, but code easily written and executed.

Two-sample *t*-test
- for Homogeneous Variances -

Begin by calculating the mean & variance for each of your two samples.

Then determine pooled variance S_p^2:

$$S_p^2 = \frac{\sum_{i=1}^{N_1}(y_i - \bar{y}_i)^2 + \sum_{j=1}^{N_2}(y_j - \bar{y}_j)^2}{(N_1 - 1) + (N_2 - 1)} = \frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{(N_1 + N_2 - 2)}$$

Theoretical formula | Machine formula

Two-sample *t*-test

- for Homogeneous Variances -

Determine the test statistic t_{calc}:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_p^2}{N_1} + \frac{S_p^2}{N_2}}} \qquad df = N_1 + N_2 - 2$$

Go to *t*-table (Appendix) at the appropriate α and df to determine t_{table}

28

Two-sample *t*-test

- Example -

Suppose a microscopist has just identified two potentially different types of cells based upon differential staining.

She separates them out in to two groups (amber cells and blue cells). She suspects there may be a difference in cell wall thickness (cwt) so she wishes to test the hypothesis:

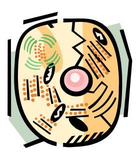
Ho: $AC_{cwt} = BC_{cwt}$
Ha: $AC_{cwt} \neq BC_{cwt}$

29

Two-sample t-test

- Example -

Parameter	AC-type	BC-type
Mean	8.57	8.40
SS	2.39	2.74
N	14	18



Notes: She counts the number of cells in one randomly chosen field of view. SS is the sum of squares (theor. formula), or numerator of the variance equation.

30

Two-sample *t*-test
- Example -

$$S_p^2 = \frac{2.39+2.74}{13+17} = 0.17$$

$$t_{calc} = \frac{8.57-8.40}{\sqrt{\frac{0.17}{14}+\frac{0.17}{18}}} = 1.13$$

$$df = 14+18-2=30$$

Ho: $AC_{cwt} = BC_{cwt}$
Ha: $AC_{cwt} \neq BC_{cwt}$

At α = 0.05/2, df = 30
$t_{table} = 2.042$

$t_{calc} < t_{table}$
Therefore Fail to reject Ho.

Cell wall thickness
is similar btw 2 types.

31

Two-sample *t*-test using R

```
> B<-c(8.8, 8.4,7.9,8.7,9.1,9.6)
> G<-c(9.9,9.0,11.1,9.6,8.7,10.4,9.5)
> var.test(B,G)

        F test to compare two variances

data:  B and G
F = 0.5063, num df = 5, denom df = 6, p-value = 0.4722
alternative hypothesis: true ratio of variances is not
equal to 1
95 percent confidence interval:
 0.0845636 3.5330199
sample estimates:
ratio of variances
           0.50633
```

32

Two-sample *t*-test using R

```
> t.test(B,G, var.equal=TRUE)

        Two Sample t-test

data:  B and G
t = -2.4765, df = 11, p-value = 0.03076
alternative hypothesis: true difference in
means is not equal to 0
95 percent confidence interval:
 -1.8752609 -0.1104534
sample estimates:
mean of x mean of y
 8.750000  9.742857
```

33

Two-sample *t*-test

- for Heterogeneous Variances -

Q. Suppose we were able to meet the normality assumption, but failed the homogeneity of variance test. Can we still perform a *t*-test?

A. Yes, but we but must calculate an adjusted degrees of freedom (df).

Two-sample *t*-test

- Adjusted df for Heterogeneous Variances -

$$df \approx \frac{\left(\frac{S_1^2}{N_1}+\frac{S_2^2}{N_2}\right)^2}{\frac{\left(\frac{S_1^2}{N_1}\right)^2}{N_1-1}+\frac{\left(\frac{S_2^2}{N_2}\right)^2}{N_2-1}}$$

Performs the *t*-test in exactly the same fashion as for homogeneous variances; but, you must enter the table at a different df. Note that this can have a big effect on decision.

Two-sample t-test using R

- Heterogeneous Variance -

```
> Captive<-c(10,11,12,11,10,11,11)
> Wild<-c(9,8,11,12,10,13,11,10,12)
> var.test(Captive,Wild)

        F test to compare two variances

data:  Captive and Wild
F = 0.1905, num df = 6, denom df = 8, p-value = 0.05827
alternative hypothesis: true ratio of variances is not
equal to 1
95 percent confidence interval:
 0.04094769 1.06659486
sample estimates:
ratio of variances
         0.1904762
```

Two-sample t-test using R
- Heterogeneous Variance -

```
> t.test(Captive, Wild)

        Welch Two Sample t-test

data:  Captive and Wild
t = 0.3239, df = 11.48, p-value = 0.7518
alternative hypothesis: true difference in means
is not equal to 0
95 percent confidence interval:
 -1.097243  1.478196
sample estimates:
mean of x mean of y
 10.85714  10.66667
```


Matched-Pair *t*-test

It is not uncommon in biology to conduct an experiment whereby each observation in a treatment sample has a matched pair in a control sample.

Thus, we have violated the assumption of independence and can not do a standard *t*-test.

The matched-pair *t*-test was developed to address this type of experimental design.

Matched-Pair *t*-test

By definition, sample sizes must be equal.

Such designs arise when:
- Same obs are exposed to 2 treatments over time.
- Before and after experiments (temporally related).
- Side-by-side experiments (spatially related).

Many early fertilizer studies used this design. One plot received fertilizer, an adjacent plot did not. Plots were replicated in a field and plant yield measured.

Matched-Pair *t*-test

The approach to this type of analysis is a bit counter intuitive.

Even though there are two samples, you will work with only one sample composed of:

STANDARD DIFFERENCES

and df = N_{ab} - 1

Matched-Pair *t*-test

- Assumptions -

The data are continuous (not discrete).

The data, i.e., the differences for the matched-pairs, follow a normal probability distribution.

The sample of pairs is a simple random sample from its population.

Matched-Pair *t*-test

$$\bar{y}_d = \frac{\sum_{d=1}^{N} y_d}{N}$$

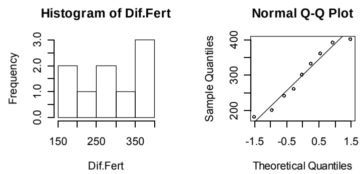
$$S_d^2 = \frac{\sum_{d=1}^{N} y_d^2 - \left(\sum_{d=1}^{N} y_d\right)^2 / N}{N-1}¿$$

$$t = \frac{\bar{y}_d}{S_d / \sqrt{N}}$$

Matched-Pair *t*-test using R

- Example -

```
> New.Fert<-c(2250,2410,2260,2200,2360,2320,2240,2300,2090)
> Old.Fert<-c(1920,2020,2060,1960,1960,2140,1980,1940,1790)

> hist(Dif.Fert)
> Dif.Fert<-(New.Fert-Old.Fert)
> shapiro.test(Dif.Fert)

Shapiro-Wilk normality
data:  Dif.Fert
W = 0.9436,
p-value = 0.6202
```



Matched-Pair *t*-test using R

- One-tailed test -

```
> t.test(New.Fert, Old.Fert,
alternative=c("greater"), mu=250, paired=TRUE)

        Paired t-test

data:  New.Fert and Old.Fert
t = 1.6948, df = 8, p-value = 0.06428
alternative hypothesis: true difference in means is
greater than 250
95 percent confidence interval:
 245.5710       Inf
sample estimates:
mean of the differences
               295.5556
```


Statistical Power

Q. What if I do a *t*-test on a pair of samples and fail to reject the null hypothesis--does this mean that there is no significant difference?

A. Maybe yes, maybe no.

Depends upon the POWER of your test and experiment.

Power

Power is the probability of rejecting the hypothesis that the means are equal when they are in fact not equal.

Power is one minus the probability of Type-II error (β).

The power of the test depends upon the sample size, the magnitudes of the variances, the alpha level, and the actual difference between the two population means.

Power

Usually you would only consider the power of a test when you failed to reject the null hypothesis.

High power is desirable (0.7 to 1.0). High power means that there is a high probability of rejecting the null hypothesis when the null hypothesis is false.

This is a critical measure of precision in hypothesis testing and needs to be considered with care.

More on Power in the next lecture.

The End

Stay tuned. Coming Soon:

Responding to Assumption Violations, Power, Two-sample Nonparametric Tests