Statistical hypothesis testing

Experimental Design and Data Analysis for Biologists

G. Quinn and M. Keough

Design by M. Logan

2004

About

This presentation is a brief revision of basic statistical concepts.

Links…

Throughout the presentation, marks in the form of (qk2002, …) provide references to sections within the recommended statistical text;

Quinn, G. P. and Keough, M. J. (2002). Experimental Design and Data Analysis for Biologists. Cambridge University Press, Cambridge.

Words and phrases in purple type face provide tooltip-style extra information, while blue type face provide links to popups that contain additional information and or definitions.

Navigation…

Navigation buttons on the right hand side of each page provide (from top to bottom) 'Previous Page', 'Next Page', 'First Page', 'Last Page', 'Go Back' and 'Quit' navigational shortcuts.

Hypothesis tests

Statistical hypothesis tests are concerned with using a sample statistic from a small subset (sample) of the total population to estimate the long run probability that a population parameter is equal to a particular value, often zero.

Null hypothesis

- A hypothesis is a prediction deduced from a model or theory
- It is impossible to prove a hypothesis, because all observations related to the hypothesis must be made
- Disproving (falsifying) is relatively straightforward
- For example, to prove that there are no foxes in Tasmania, it is necessary to survey the entire state simultaneously. On the other hand to disprove the notion that there are no foxes in Tasmania, only one fox needs to be found

- Hence, it is usual to test (attempt to falsify) a null hypothesis (H_0) – which includes all possibilities except the prediction in the hypothesis
- If the hypothesis of interest is that there **is** an effect (difference, relationship ...), then the H_0 would be that there is **no** effect.
- Thus disproving the H_0 will provide evidence that the actual hypothesis is true.

Simple null hypothesis

- **Test of hypothesis that a population mean or combination of means equals a particular value (H_0: $\mu = \theta$).**
  - e.g. H_0 that the population mean SO_4^{2-} concentration in forested North American streams is 60 ($\theta = 60$) Lovett *et al*. (2000).
  - e.g. H_0 that the population mean density of kelp after a potential impact (Nuclear power plant) is the same as before ($\theta = 0$)
- **These values may be from the literature or other research or set by legislation**

Hypothesis tests and the *t* statistic

- The simplest H_0 test is where we test whether a population mean (or difference between two means) equals a certain value (θ).
- The basis of hypothesis testing is to generate a statistic from the sample data, and compare this statistic to its known probability distribution under the H_0.
- As the mean of any given sample usually does not have a known probability distribution to which it can be compared, it is usual to generate a test statistic (*t* statistic) that does have a known probability distribution (*t*-distribution).

Hypothesis tests and the *t* statistic

- The general form of a *t* statistic is:

$$t = \frac{S_t - \theta}{SE}$$

where S_t is the sample statistic, θ is the parameter value specified in H_0 and SE is the standard error of the statistic.

- The specific form for testing means is:

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

value of mean specified in H_0, i.e. θ

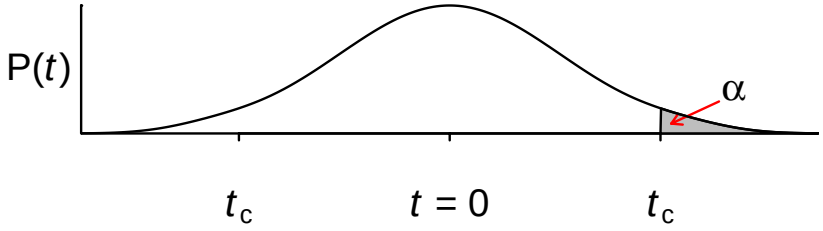
Statistical hypothesis testing

- Statistical null hypothesis (H_0)
  - hypothesis of no difference (or no relationship, or no effect, etc)
- H_0 refers to population parameters:
  - e.g. no difference between population means or no correlation in the population
- If H_0 is false (rejected), then some H_A (alternative hypothesis, or working hypothesis) must be true

Test statistics

- For each possible sample size ($df = n-1$), there is a unique sampling distribution of t when the H_0 is true
- The area under each sampling (probability) distribution is one
- Hence it is possible to determine the probability of obtaining particular ranges of t values when the H_0 is true
  - e.g. the probability of obtaining a particular t value or greater if the H_0 is true

Sampling distribution of *t*



	Critical α			
df	**0.10**	**0.05**	**0.01**	**0.001**
1	3.078	6.314	31..821	318.309
2	1.886	2.920	6.965	22.327
3	1.638	2.353	4.541	10.215
20	1.325	1.725	2.528	3.552
∞	1.282	1.645	2.324	3.090

Expurgated table of critical *t*-values. For each *df* and α there is a critical *t*-value (t_c). If your sample *t*-value is greater or equal to t_c, reject the H_0 (significant result)

Decision criteria

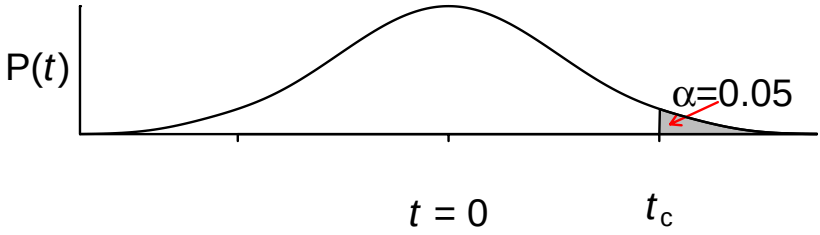
- How low a probability should make us reject our H_0?
  - the probability of obtaining our test statistic or one greater, if the H_0 is true
- If the probability is less than the predetermined significance level (α), then reject H_0; otherwise do not reject.
- Significance level conventionally set to $\alpha = 0.05$ (5%)
- α is arbitrary
  - other significance levels might be valid
  - trade-off between type I and type II errors (QK2002….)

One tailed tests

$H_0: \mu \leq \theta$ $H_A: \mu > \theta$

$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$

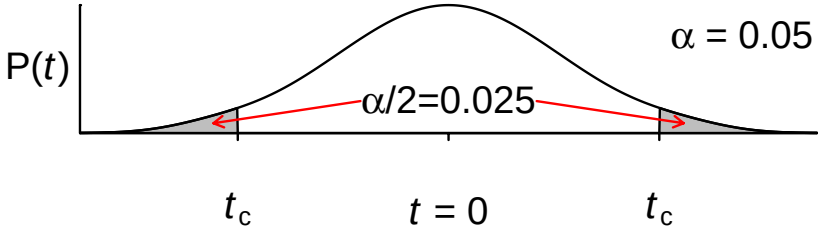
- So only reject H_0 for large +ve values of t, i.e. when sample mean is much greater than θ.



Two tailed tests

H_0: $\mu = \theta$ H_A: $\mu > \theta$ or H_A: $\mu < \theta$ $\quad t = \frac{\bar{y}-\mu}{s/\sqrt{n}}$

- So reject H_0 for large +ve or –ve values of t, i.e. when sample mean is much greater or less than θ.



t-tests

Single population

- H_0: $\mu = 0$ (or any other pre-specified value)

$$t = \frac{\bar{y} - 0}{S_{\bar{y}}} = \frac{\bar{y} - 0}{s/\sqrt{n}}$$

df = *n* -1

t tests

Two populations -with independent observations

- H_0: $\mu_1 = \mu_2$, i.e. $\mu_1 - \mu_2 = 0$

$$t = \frac{\bar{y}_1 - \bar{y}_2 - (\mu_1 - \mu_2)}{s_{\bar{y}_1} - s_{\bar{y}_2}} = \frac{\bar{y}_1 - \bar{y}_2}{s_{y_1} - y_2}$$

$df = (n_1 - 1) + (n_2 - 1) = n_1 + n_2 - 2$

Examples of *t* tests on two populations

- **H_0 no difference between zones in the mean number of eggs per capsule of intertidal predatory snail (Ward and Quinn, 1988).** **(qk2002, Box 3.1)**

- **H_0 no difference between sexes in the mean metabolic rates of fulmars, a species of seabird (Furness and Bryant, 1996).** **(qk2002, Box 3.2)**

Paired *t* tests

Paired observations

H_0: $\mu_d = 0$

where d is the difference between paired observations

$$t = \frac{\bar{d}}{s_{\bar{d}}}$$

Where $\bar{d}$ is the mean of the pairwise differences and $S_{\bar{d}}$ is the standard error of the pairwise differences.

$df = (n - 1)$

Example of paired *t* tests

- **H_0 no difference in size of webs of orb-spinning spiders in light compared to dark.**
  - same spiders were used in both light regimes (Elgar et al., 1996). (qk2002, Box 3.2)

Testing a statistical null hypothesis

Worked example

- **Fecundity of predatory gastropods: (Ward and Quinn, 1988).** **(qk2002, Box 3.1)**
  - collected 37 and 42 gastropod egg capsules from the littorinid (high on shore with littorinid snails) and mussel zone (mid shore with mussel beds) respectively
- **Counted the number of eggs per capsule**
- **Null hypothesis (H_0):**
  - no difference between zones in the mean number of eggs per capsule

Hypothesis testing – Step 1

- **Specify H_0 and select appropriate test statistic:**
  - H_0: $\mu_L = \mu_B$, i.e. the population mean number of eggs per capsule is the same for each zone

The *t* statistic is an appropriate test statistic for comparing two population means

Hypothesis testing – Step 2

- **Specify the *a priori* significance (probability) level (α)**

By convention, use $\alpha = 0.05$ (5%)

Hypothesis testing – Step 3

- **Conduct experiment, satisfy test assumptions (see Est.pdf and Eda.pdf) and calculate the test statistic from sample data:**

	Mean	SD	*n*
Littorinid	8.70	3.03	37
Mussel	11.36	2.33	42

***t* = -5.39, *df* = 77**

Note: the sign is not important (only the magnitude) since a *t* distribution is symmetrical about 0

Hypothesis testing – Step 4

- Compare the value of *t* statistic to its sampling distribution, the probability distribution of the statistic (for the specific *df*) when H_0 is true:
  - what is the probability of obtaining a *t* value of 5.39 or greater when the H_0 is true?
  - what is the probability of obtaining a *t* value of 5.39 or greater from a *t* distribution with 77*df* ?
  - what is the probability of collecting samples with the observed degree of difference between group means (or one greater) from two populations with the same means?

Hypothesis testing – Step 5

- *P* value can be obtained from statistical software
  - in this example, $P = 0.000000746$
- Critical *P* value for the specific df can be obtained from a *t* lookup table

Hypothesis testing – Step 6

- If the probability of obtaining our *t* value (5.39) or one greater is less than α, conclude that H_0 is "unlikely" to be true and reject it:
  - statistically significant result
- Our probability (0.000000746) is much less than 0.05, therefore it is very unlikely that our samples came from populations that had the same mean number of eggs per capsule.
- Hence, the H_0 is rejected (statistically significant result)

Hypothesis testing – Step 6 cont.

- If the probability of obtaining our *t* value or one greater is greater or equal to α, conclude that H_0 is **not** "unlikely" to be true and **do not** reject it:
  - statistically non-significant result

Presenting *t* tests in reports

Methods section

An independent *t* test was used to compare the mean number of gastropod eggs per capsule from Littorinid and Mussel zones. Assumptions were checked ...

Results section

The mean number of eggs per capsule from the Mussel zone was found to be significantly higher than that of the Littorinid zone ($t = 5.39$, df = 77, $P < 0.001$; see Fig. 2)

Assumptions of a *t* test

- All statistical tests (not just *t* tests) have assumptions
- *P* values from statistical tests are only reliable when these assumptions are not violated
  - assumptions must be met so test statistic from sample can be compared to probability distribution for that statistic
- All statistical tests (*t* tests included) assume that all observations are independent or specifically paired (independence assumption)

Assumptions of a *t* test

- The *t* test is a parametric test
  - it assumes a specific distribution of the response variable
  - most parametric tests assume a normal distribution
- The *t* statistic only follows a *t* distribution if:
  - the populations from which the samples were collected is normally distributed (normality assumption)
  - the two populations have equal variances (homogeneity of variance assumption)

Independence assumption

- Samples are collected to estimate and represent the total definable population
- To ensure that a sample will be a representative of the population (without bias), each observation needs to be independent
- Bias is minimized through random sampling (or at least haphazard

Normality assumption

- The data in each group are normally distributed

- Checks/diagnostics: (also see Eda.pdf)
  - frequency distributions
  - boxplots
  - formal tests for normality (however, realistic sample sizes rarely allow such tests)

Normality assumption

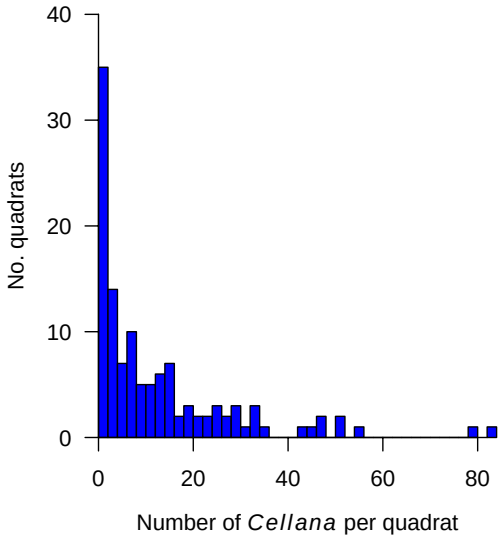
- Often non normal data can be normalised using a simple transformation such as a logarithmic or square root transformation (see Eda.pdf)
- However, if data can not be normalised, then:
  - non-parametric analyses should be explored

Homogeneity of variances assumption

- The population variances of the two groups are equal

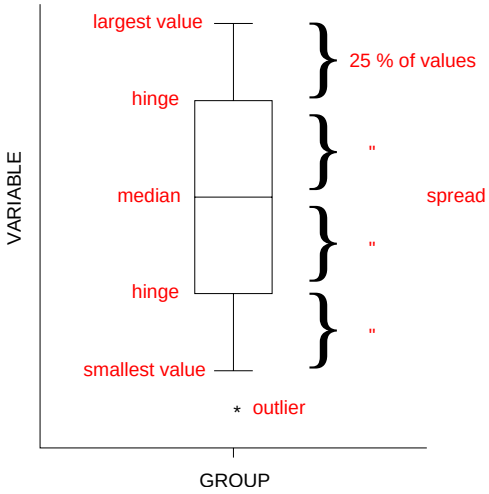
- Checks/diagnostics: (also see Eda.pdf)
  - subjective comparison of sample variances
  - boxplots
  - *F*-ratio test of H_0: $\sigma_1^2 = \sigma_2^2$

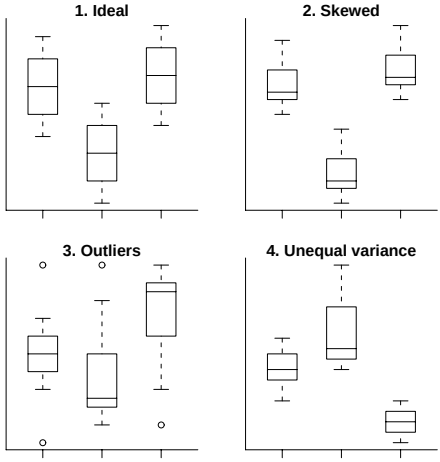
Frequency histograms



This graph represents a skewed distribution. The number of limpets per quadrat are not normally distributed.

Boxplots





Nonparametric tests

- Usually based on ranks of the data
- H_0: samples come from populations with identical distributions
- H_A: only difference is in location (medians)
- Don't assume particular underlying distribution of data
  - normal distributions not necessary
- Equal variances and independence still required

Mann-Whitney-Wilcoxon test

- Calculates the sum of the ranks in 2 samples
  - should be the same if H_0 is true
- Compare the rank sum to its H_0 sampling distribution
  - the distribution of rank sums when H_0 is true
- Equivalent to a *t* test calculated on data transformed to ranks

References

- Elgar, M. A., Allan, R. A., and Evans, T. A. (1996). Foraging strategies in orb-spinning spiders: ambient light and silk decorations in *Argiope aetherea* Walckeraer (Araneae: Araneoidea). *Australian Journal of Ecology*. **21**:464–467.

- Furness, R.W. and Bryant, D. M. (1996). Effect of wind on field metabolic rates of breeding Northern Fulmars. *Ecology*. **77**:1181–1188.

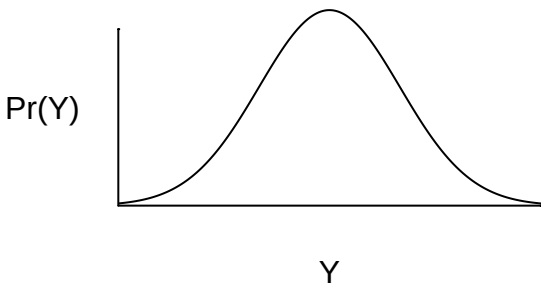
- Lovett, G. M., Weathers, K. C., and Sobczak, W. V. (2000). Nitrogen saturation and retention in forested watersheds of the Catskill Mountains, New York. *Ecological Applications*. **10**:73–84.

- Ward, S. and Quinn, G. P. (1988). Preliminary investigations of the ecology of the predatory gastropod *Lepsiella vinosa* (Lemarch) (Gastropoda Muricidae). *Journal of Molluscan Studies*. **54**:109–117.

Normal distribution

The normal (or Gaussian) distribution is a symmetrical probability distribution with a characteristic bell-shape.

The standard normal distribution (z distribution) is a normal distribution with a mean of zero and a standard deviation of one.



The normal distribution is the most important probability distribution for data analysis; most commonly used statistical procedures in biology assume that the variables follow a normal distribution.

Click anywhere to close

Parametric vs non-parametric analyses

Parametric tests make distributional assumptions about the population(s) from which the data were sampled. Hence, they can only be applied to data from which the probability distribution(s) for the sampled population(s) can be specified. As an example, statistical tests that are based on the *t* distribution assume that the populations from which the samples were collected are normal. Contrastingly, **nonparametric** rank based tests do not make any distributional assumptions since they generate their own probability distribution for the particular test statistic. Observations are ranked and the ranks are then randomized a large number of times (each time recalculating the test statistic) to generate a probability distribution of the rank-based test statistic.

Type I and type II errors

A Type I error is falsely rejecting the H_0. That is rejecting the H_0 (obtaining a statistically significant result) when the H_0 is actually true.

- To reduce the risks of Type I errors, the decision criterion (α) is set to as low as possible

A Type II error is falsely accepting the H_0. That is accepting the H_0 (obtaining a statistically non-significant result) when the H_0 is actually false.

- To reduce the risks of Type II errors, the decision criterion (α) is increased
- Hence, there is a compromize between Type I error and Type II error risk minimisation, and thus by convention, α is set to 0.05.
- Lowering the α to reduce the risks of a type I error will increase the risk of a Type II error.

Click anywhere to close

Population parameters vs sample statistics

A **population** refers to all the possible observations. Therefore, population parameters refer to the characteristics of the whole population. For example, the population mean.

A **sample** represents a collected subset of the population's observations and is used to represent the entire population. Sample statistics are the characteristics of the sample (e.g. sample mean) and are used to estimate population parameters.

Click anywhere to close

Degrees of freedom (*df*)

The degrees of freedom is the number of observations in our sample that are 'free to vary' (not fixed) when we are estimating the variance.

To calculate the variance of a sample, the mean of a sample must be known, and therefore, only n_1 observations are free to vary.

If we know the mean and the value of n_1 observations, then the value of the 1 remaining observation must be known (fixed) and is thus not free to vary. (qk2002, Box 2.1)

Click anywhere to close