Statistical Geophysics

Chapter 3

Inferential Statistics 2

Inferential Statistics 2

Introduction to hypothesis testing

Background

- So far, we have considered problems of estimation.
- We will now study what are generally called tests of hypotheses.
- These tests yield a binary decision that a particular hypothesis about a phenomenon generating the data may be true or not.
- There are two types of tests: parametric tests and nonparametric (or distribution-free) tests.

Sampling distribution

- The sampling distribution for a statistic (including the test statistic for a hypothesis test) is the probability distribution describing batch-to-batch variations of that statistic.

- The value of a statistic computed from a particular batch of data will in general be different from that for the same statistic computed using a different batch of data of the same kind.

- Example: Average January temperature is obtained by averaging daily temperatures during that month at a particular location for a given year. The statistic is different from year to year.

Elements of any hypothesis test

1. Identify a test statistic that is appropriate to the data and question at hand.
2. Define a null hypothesis, H_0, which defines a reference against which to judge the observed test statistic.
3. Define an alternative hypothesis, H_1 (or H_A).
4. Obtain the null distribution, which is the sampling distribution for the test statistic, if H_0 is true.
5. Compare the observed test statistic to the null distribution. If the test statistic falls in a sufficiently improbable region of the null distribution, H_0 is rejected as too implausible to have been true given the observed evidence.

Test level

- The sufficiently improbably region of the null distribution is defined by the rejection level (or test level) of the test.
- H_0 is rejected if the probability of the observed test statistic, and all other results at least as unfavourable to H_0, is less than or equal to the test level.
- The test level is chosen in advance of the computations.
- Commonly the 5% level is chosen, although tests conducted at the 10% level or the 1% level are not unusual.

p value

- The p value is the probability that the observed value of the test statistic, together with all other possible values of the test statistic that are at least as unfavourable to H_0, will occur.

- Thus, H_0 is rejected if the p value is less than or equal to the test level and is not rejected otherwise.

- The p value also communicates the confidence with which a null hypothesis has or has not been rejected.

Error types and power of a test

	H_0 is true	H_0 is false
H_0 is rejected	Type I error	No error
H_0 is not rejected	No error	Type II error

We define

$$\begin{aligned}\alpha &= \mathrm{P}(\text{type I error}) = \mathrm{P}(\text{reject } H_0 | H_0 \text{ true}) \\ \beta &= \mathrm{P}(\text{type II error}) = \mathrm{P}(\text{not reject } H_0 | H_0 \text{ false}) \ .\end{aligned}$$

The quantity $1 - \beta$ is known as the power of a test against a specific alternative.

Error types and power of a test

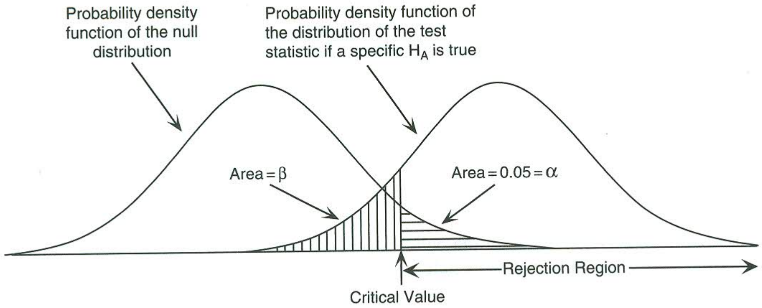

Figure: Illustration of the relationship of the probability of a Type I error (horizontal hatching) and the probability of a Type II error (vertical hatching) for a test conducted at the 5% level.

One-sided versus two-sided tests

- A statistical test can be either one-sided or two-sided.
- A one-sided test is appropriate
  - ...if there is a prior reason to expect that violations of H_0 will lead to values of the test statistic on a particular side of the null distribution.
  - ...when only values on one tail or the other of the null distribution are unfavorable to H_0, because the way the test statistic has been constructed.
- Two-sided tests are appropriate when either very large or very small values of the test statistic are unfavourable to the null distribution.

Confidence intervals: inverting hypothesis tests

- There is a duality between a one-sample hypothesis test and the computed confidence interval (CI) around the observed statistic.

- The $100 \times (1 - \alpha)\%$ CI around an observed statistic
  - will not contain the null hypothesis value of the test if the test is significant at the α level,
  - and will contain the null value if the test is not significant at the α level.

Example: Exact binomial test

- Advertisements for a tourist resort claim that, on average, six days out of seven are cloudless during winter ($6/7 = 0.857$).
- Assume that we could arrange to take observations on 25 independent occasions.
- If cloudless skies are observed on 15 of those 25 days, is this observation consistent with, or does it justify questioning, the claim?
- This problem fits neatly into the parametric setting of the binomial distribution.

Example: Exact binomial test

- The test statistic of $X = 15$ out of $n = 25$ days has been dictated by the form of the problem.

- Test problem: H_0: $\pi \geq 0.857$ vs. H_1: $\pi < 0.857$.

- For this test, the null distribution is binomial, with parameters $n = 25$ and $\pi = 0.857$.

- The p value of this exact binomial test is

$$P(X \leq 15) = \sum_{x=0}^{15} \binom{25}{x} 0.857^x (1 - 0.857)^{25-x} = 0.0015 \; .$$

Example: Exact binomial test

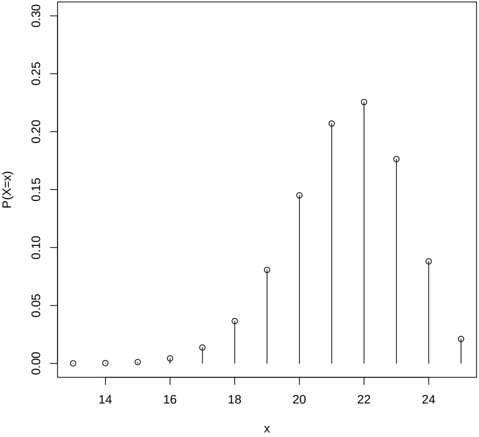

Figure: Exact binomial null distribution $\mathcal{B}(n = 25, \pi = 0.857)$.

Approximate binomial test

Approximation of the binomial distribution:

- Let $X = \sum_{i=1}^{n} X_i \sim \mathcal{B}(n, \pi)$. It follows from the Central Limit Theorem that for sufficiently large n:

$$X \overset{a}{\sim} \mathcal{N}(n\pi, n\pi(1-\pi))$$

and

$$Z = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \overset{a}{\sim} \mathcal{N}(0, 1) \quad .$$

Approximate binomial test

Summary

- Suppose the following test problems for the parameter π of the $\mathcal{B}(n, \pi)$ distribution:
  (a) $H_0 : \pi = \pi_0$ vs. $H_1 : \pi \neq \pi_0$
  (b) $H_0 : \pi \geq \pi_0$ vs. $H_1 : \pi < \pi_0$
  (c) $H_0 : \pi \leq \pi_0$ vs. $H_1 : \pi > \pi_0$.

- Based on the observed test statistic

$$z = \frac{x - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

  and given α, H_0 is rejected if
  (a) $|z| > z_{1-\alpha/2}$
  (b) $z < -z_{1-\alpha}$
  (c) $z > z_{1-\alpha}$.

Example: Approximate binomial test

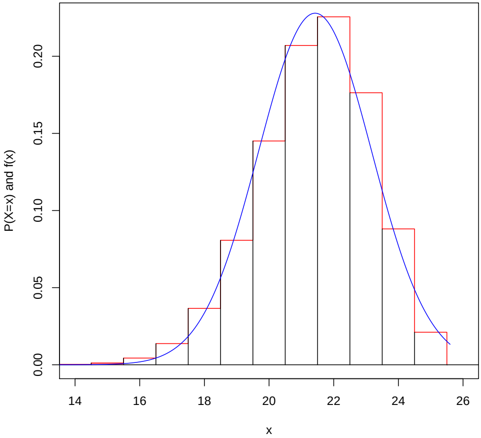

Figure: Relationship of the binomial null distribution (histogram bars), and its Gaussian approximation (smooth curve).

Continuity correction

Approximation of the binomial with continuity correction:

- Let $X \sim \mathcal{B}(n, \pi)$. For sufficiently large n:

$$\begin{aligned} \mathrm{P}(X \leq x) &\approx \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \\ \mathrm{P}(X = x) &\approx \Phi\left(\frac{x + 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) - \Phi\left(\frac{x - 0.5 - n\pi}{\sqrt{n\pi(1-\pi)}}\right) \quad . \end{aligned}$$

Inferential Statistics 2

Some commonly encountered parametric tests

t test for the mean

- We want to compare a hypothetical mean, μ_0, with the true unknown mean μ.

- If the number of data values making up the sample mean is large enough for its sampling distribution to be essentially Gaussian, then the test statistic

$$T = \frac{\bar{X} - \mu_0}{S}\sqrt{n} \ ,$$

  where $S = \sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)}$, follows a *t* distribution with $n-1$ degrees of freedom (d.f.).

- The statistic T resembles the standard Gaussian variable $Z = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$, except that a sample estimate of the variance of the sample mean has been substituted in the denominator.

t test for the mean

Summary

- Consider the test problems:

  (a) $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
  (b) $H_0 : \mu \geq \mu_0$ vs. $H_1 : \mu < \mu_0$
  (c) $H_0 : \mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$.

- Based on the observed test statistic

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

  and given α, H_0 is rejected if

  (a) $|t| > t_{1-\alpha/2}(n-1)$
  (b) $t < t_{\alpha}(n-1) = -t_{1-\alpha}(n-1)$
  (c) $t > t_{1-\alpha}(n-1)$.

- For $n \geq 30$: Approximate quantiles of the t distribution of $n-1$ d.f. by quantiles of the $\mathcal{N}(0,1)$ distribution.

Comparison of two proportions

- Suppose the data are categorized into two groups.
- Let π_1 (π_2) be the probability of success in group 1 (group 2).
- Test problem H_0: $\pi_1 = \pi_2$ vs. H_1: $\pi_1 \neq \pi_2$.
- Sample is presented as a two-by-two contingency table:

	Group 1	Group 2	$\sum$
Success	10	15	25
Failure	20	15	35
$\sum$	30	30	60

- Proportions of success:

  Group 1: $\frac{10}{30} = 33\% = \hat{\pi}_1$, Group 2: $\frac{15}{30} = 50\% = \hat{\pi}_2$.

Test statistic

- The test statistic is

$$\chi^2 = \sum_{i=1}^{4} \frac{(o_i - e_i)^2}{e_i} \quad ,$$

  where $i = 1, \ldots, 4$ are the four cells in the middle of the contingency table.

- The o_i are the observed counts and the e_i are what is expected if $\pi_1 = \pi_2$.

Computing e_i

- If H_0: $\pi_1 = \pi_2$ is true, we could estimate the common probability π by $\hat{\pi} = 25/60 = 0.4167$.
- In the upper left corner we would expect to see $0.4167 \times 30 = 12.501$ successes in group 1, and so $30 - 12.501 = 17.499$ failures in the lower left.
- In the upper right corner we would expect to see $0.4167 \times 30 = 12.501$ successes in group 2, and so $30 - 12.501 = 17.499$ failures in the lower right.

Decision

- The value of the observed test statistic χ^2 is

$$\chi^2 = \frac{(10-12.501)^2}{12.501} + \frac{(20-17.499)^2}{17.499} + \frac{(15-12.501)^2}{12.501} + \frac{(15-17.499)^2}{17.499} = 1.7142 \ .$$

- The sampling distribution of this test statistic is the χ^2 distribution with 1 degrees of freedom (d.f.).

- Reject H_0 if $\chi^2 > \chi^2_{1-\alpha}(1)$.

- $\chi^2 = 1.7142 < \chi^2_{0.95}(1) = 3.8415$. The observed difference is statistically not significant at the 5% level (p value = 0.1905).

Inferential Statistics 2

Test for differences of mean under independence

Example: comparison of two fertilizers

- Response: crop yield.
- Two independent samples (each with sample size $n = 6$).
- Crop yield using
  - fertilizer X: $22, 21, 18, 16, 22, 17$.
  - fertilizer Y: $20, 22, 17, 13, 17, 18$.
- μ_X (μ_Y) denotes the mean crop yield using fertilizer X (Y).
- Test problem: H_0: $\mu_X = \mu_Y$ vs. $H_1 : \mu_X \neq \mu_Y$.

Two-sample t test

- $X_k \overset{i.i.d.}{\sim} \mathcal{N}(\mu_X, \sigma_X^2)$ $(k = 1, \ldots, n)$ and $Y_l \overset{i.i.d.}{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$ $(l = 1, \ldots, m)$.

- Assumption: $\sigma_X = \sigma_Y$ (unknown).

- Test problem: $H_0 : \mu_X - \mu_Y = \delta_0$ vs. $H_1 : \mu_X - \mu_Y \neq \delta_0$.

- Observed test statistic:

$$t = \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\left\{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}\right\}}} \;.$$

- Reject H_0, if $|t| > t_{1-\frac{\alpha}{2}}(n + m - 2)$.

Example:

- The sample means are

$$\bar{x} = \frac{1}{6}(22 + 21 + 18 + 16 + 22 + 17) = 19.33$$

$$\bar{y} = \frac{1}{6}(20 + 22 + 17 + 13 + 17 + 18) = 17.83$$

- The observed difference is $\bar{x} - \bar{y} = 1.5$.

- An estimate for the pooled sample variance is 8.2167.

- The value of the observed test statistic is $t = \frac{1.5}{\sqrt{\frac{1}{3} \times 8.2167}} = 0.9064$.

- Decision: $t = 0.9064 < t_{0.975}(10) = 2.2814$: H_0 is not rejected (p value = 0.386).

Inferential Statistics 2

Test for differences of mean for paired samples

Paired t test

- An important from of non-independence occurs when the data values are paired.
- The two-sample t test for paired data analyzes the differences $D_i = X_i - Y_i$ $(i = 1, \ldots, n)$ between corresponding members of the $n_1 = n_2 = n$ pairs.
- The test statistic is

$$T = \frac{\bar{D} - \mu_D}{S_D}\sqrt{n} \ ,$$

where $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$, $\mu_D = \mu_X - \mu_Y$ and $S_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2}$.

- The test problem is transformed to the one-sample setting.