24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

24.910

Laboratory Phonology

Basic statistics

Reading:

- Fowler & Housum 1987.

Assignments:

- Write up voicing perception experiment (due in two weeks 5/8).
- Progress report on your project (5/1).
- Project draft + presentation 5/15.

Topics:

- Statistics
- The lexicon and context in speech perception.
- The lexicon and context in speech production.
- Phonology in speech perception.

Writing up an experiment

The report on an experiment usually consists of four basic parts:

1. Introduction
2. Procedure
3. Results
4. Discussion

Writing up an experiment

1. Introduction
- Outline of the purpose of the experiment
- state hypotheses tested etc
- provide background information (possibly including descriptions of relevant previous results, theoretical issues etc).

2. Procedure - what was done and how.
- instructions for replication, e.g.
  - Experimental materials
  - Subjects
  - Recording procedure
  - Measurement procedures (especially measurement criteria**).**

Writing up an experiment

3. Results

- Presentation of results, including descriptive statistics (means etc) and statistical tests of hypotheses.

4. Discussion

- Discuss the interpretation and significance of the results

Some Statistics

Two uses of statistics in experiments:

- Summarize properties of the results (descriptive statistics).
- Test the significance of results (hypothesis testing).

Descriptive statistics

A measure of central tendency:

- Mean:

$$M = \frac{\Sigma x_i}{N}$$

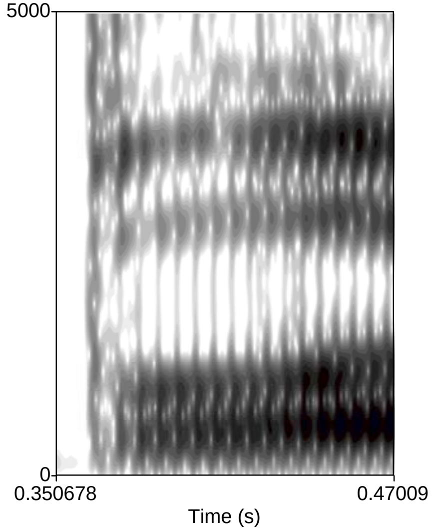
  - M is used for sample mean, μ for population mean.

A measure of dispersion:

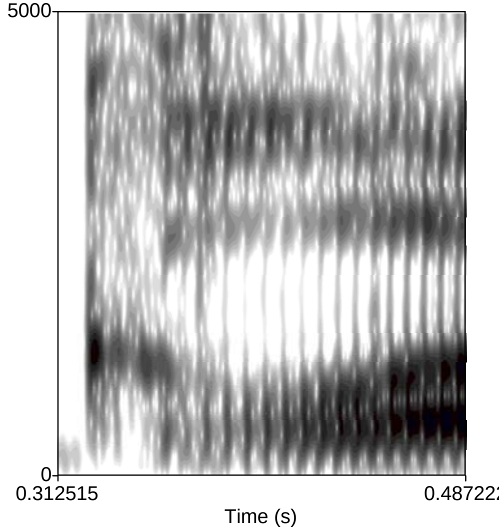
- Variance: mean of the squared deviations from the mean

$$\sigma^2 = \frac{\Sigma(x_i-\mu)^2}{N}$$

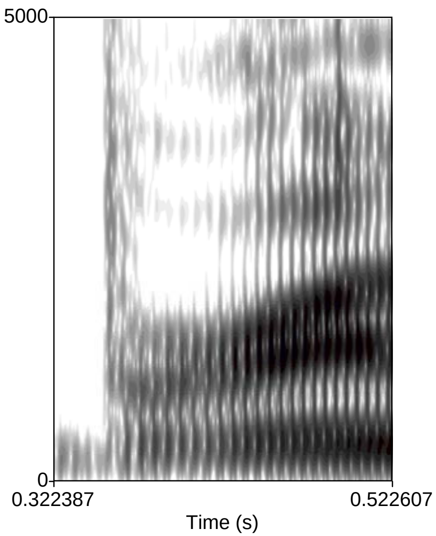
- Standard deviation: σ (square root of the variance).



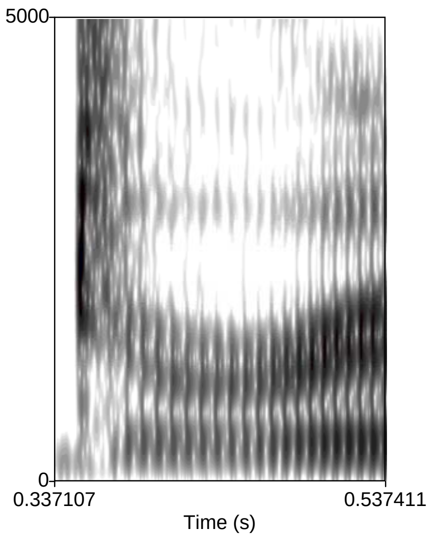
bl(ow)



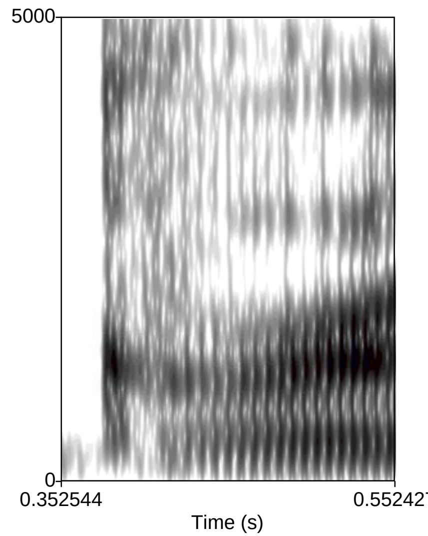
gl(ow)



br(ew)



dr(ew)



gr(ew)

Hypothesis Testing

- F2 onset (Hz)

	br	dr	gr
mean	1225	1641	1272
s.d.	150	177	215

- Are these differences in means significant?
- Could the apparent differences have arisen by chance, although the true (population) means of F2 onsets are the same?

- I.e. given that F2 onsets vary, we might happen to sample most of our [br] onsets from the low end of the distribution, and most of our [gr] onsets from the high end.
- Statistical tests allow us to assess the probability that this is the case.

Hypothesis Testing: *t*-test

- The *t*-test allows us to test hypotheses concerning means and differences between means.
  1. 'Mean F2 onset in [br] differs from mean F2 onset in [gr] in English'.
  2. 'Mean F2 onset in [br] is 1250 Hz' (unlikely, but a simpler case - cf. [afva] is identified as [afa] > 50%).

- We actually evaluate two exhaustive and mutually exclusive hypotheses, a **null hypothesis** that the mean has a particular value, and the alternative hypothesis that the mean does not have that value.
  1. The mean F2 onset in [br] is the same as the mean F2 onset in [gr] (Null).
  2. The mean F2 onset in [bt] $\neq$ 1250 Hz (Alternative).

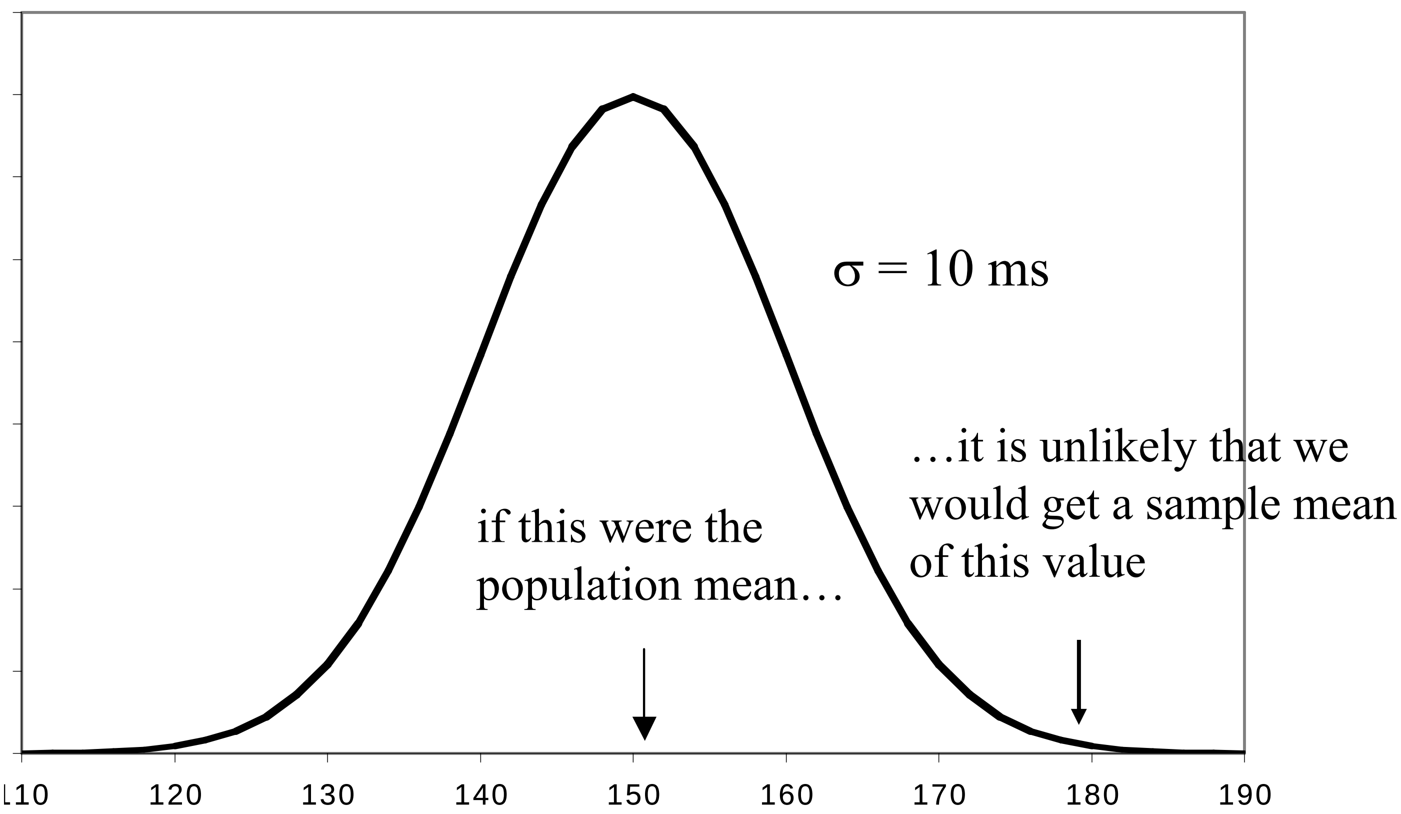
- Statistical tests allow us to assess the probability of obtaining the observed data if the null hypothesis were true.

Hypothesis Testing: *t*-test

- Basic concept: If we know what the distribution of sample means would be if the null hypothesis were true, then we can calculate the probability of obtaining the observed mean, given the null hypothesis.
- We arrive at the parameters of the distribution of sample means through assumptions and estimation.

Hypothesis Testing

Distribution of sample means



Hypothesis Testing

- Basic assumption: The samples are drawn from normal populations.

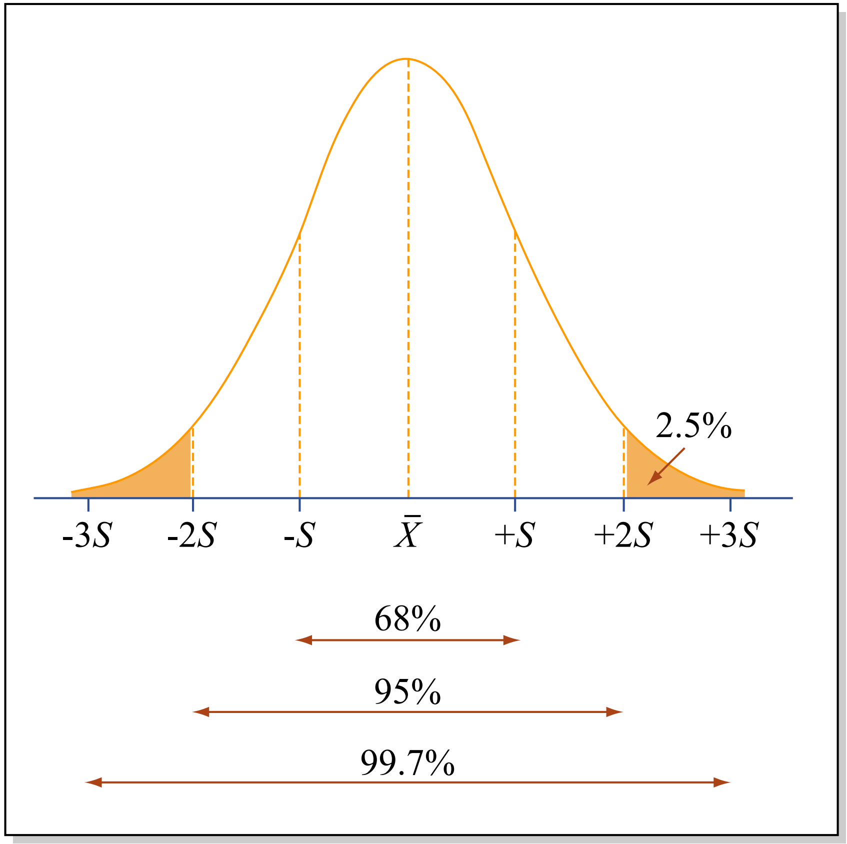


Figure by MIT OpenCourseWare. Adapted from Kachigan, S. K. *Multivariate Statistical Analysis*. 2nd ed. New York, NY: Radius, 1991.

Hypothesis Testing

- Basic assumption: The samples are drawn from normal populations.
- Properties of distribution of means of samples of size *N* drawn from a normal population:
  - The sample means are normally distributed.
  - Mean is the same as the population mean.
  - The variance is less than the population variance:

$$\sigma_M^2 = \frac{\sigma^2}{N}$$

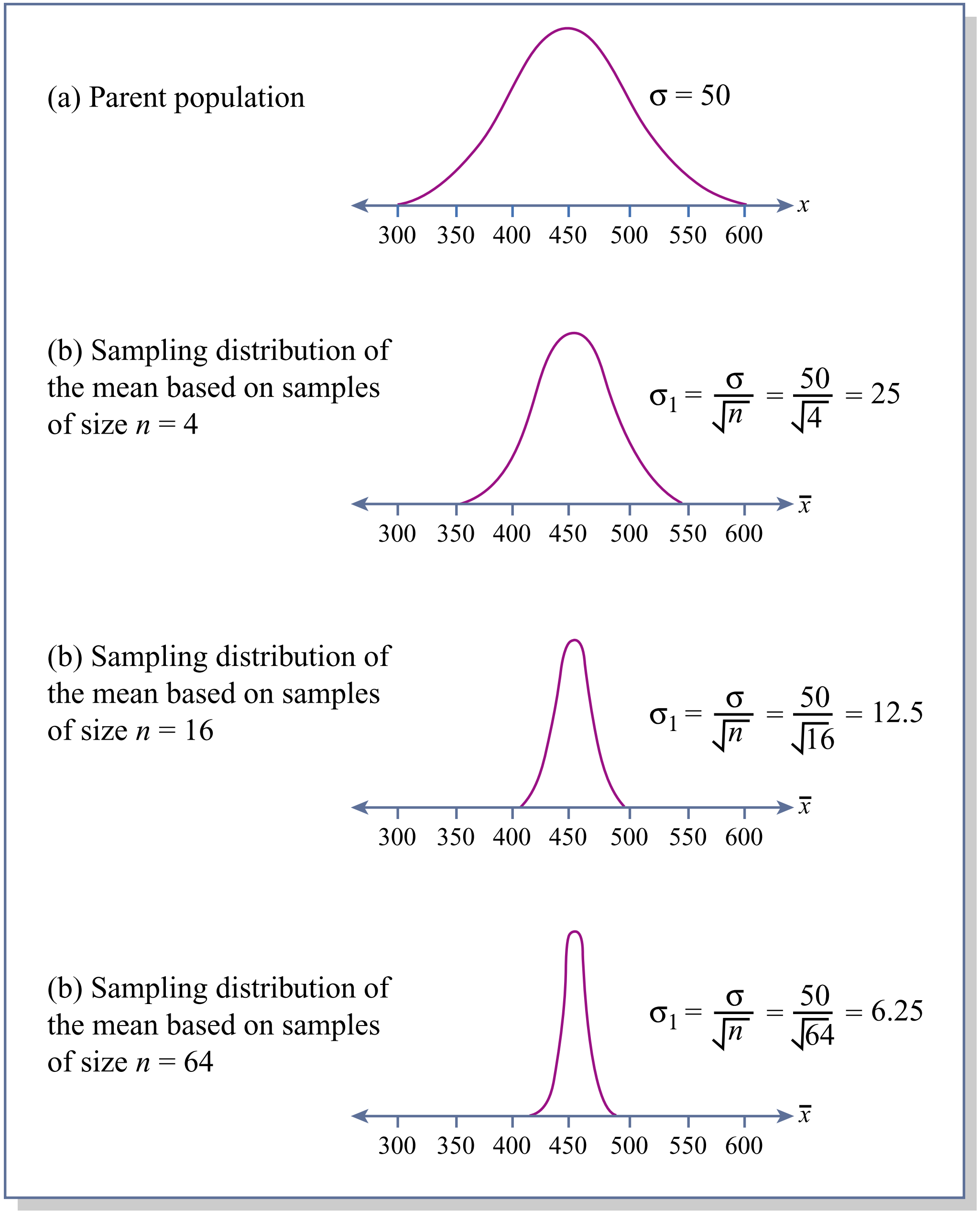


Figure by MIT OpenCourseWare. Adapted from Kachigan, S. K *Multivariate Statistical Analysis*. 2nd ed. New York, NY: Radius, 1991.

Hypothesis Testing

- The mean of the distribution is determined by hypothesis.
  - E.g. mean = 1250 Hz or mean difference = 0.
- Population variance is estimated from the sample variance. Unbiased estimate of the population variance:

$$S^2 = \frac{\Sigma(x_i - M)^2}{N-1}$$

  - N-1 is the number of degrees of freedom of the sample.
- So estimated variance of distribution of sample means, $S_M^2 = S^2/N$
- *t* score:

$$t = \frac{M - \mu}{S_M}$$

Hypothesis Testing

- *t* scores follow a *t*-distribution - similar to a normal distribution, but with slightly fatter tails (more extreme values) because S may underestimate σ.
- *t*-distribution is actually a family of distributions, one for each number of degrees of freedom.
- Calculate *t*-score then consult relevant *t* distribution to determine the probability of obtaining that *t*-score or greater (more extreme).

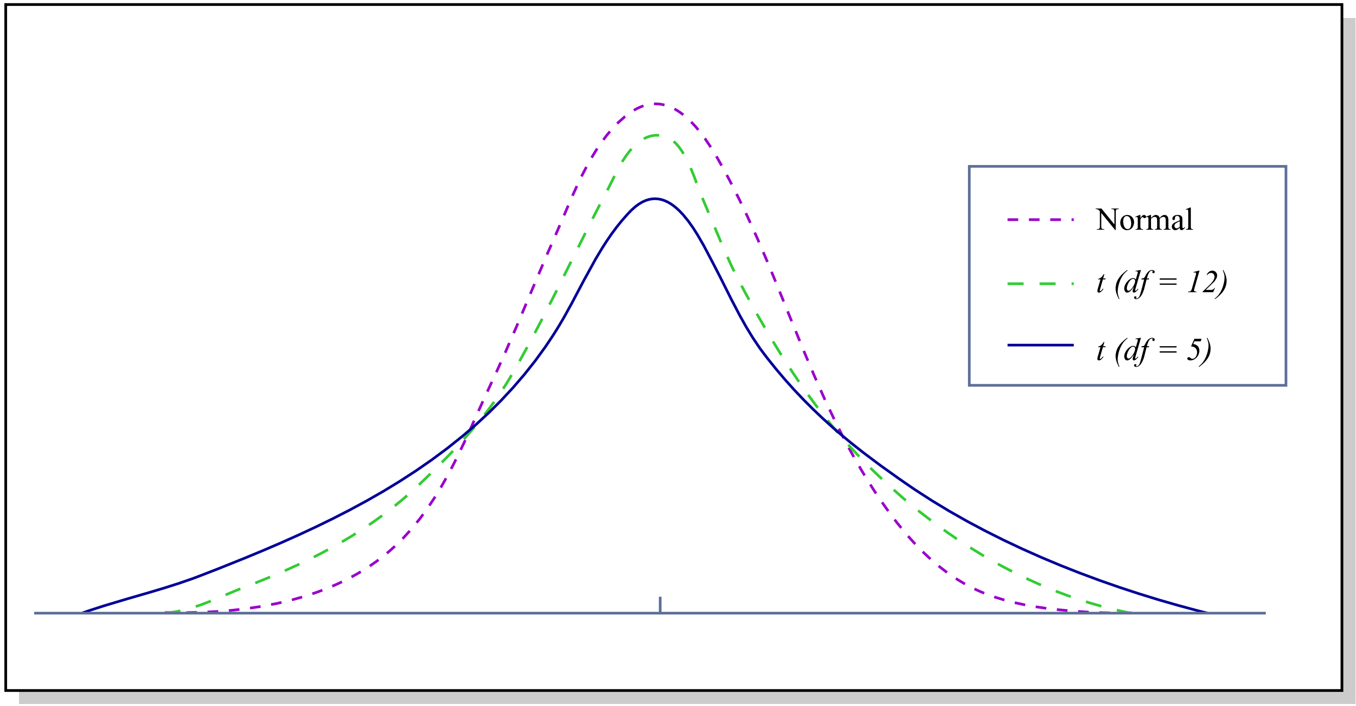


Figure by MIT OpenCourseWare.

t test for independent means

- When we compare means, we are actually sampling a population of differences (e.g. differences in durations of vowels in open and closed syllables).
- If the null hypothesis is correct, then the mean difference is 0.
- Variance of the distribution of mean differences is estimated based on the variances of the two samples.

Hypothesis testing

- Statistical tests like the *t* test give us the probability of obtaining the observed results if the null hypothesis were correct - the ‘p’ value. E.g. $p < 0.01$, $p = 0.334$.
- We reject the null hypothesis if the experimental results would be very unlikely to have arisen if the null hypothesis were true.
- How should we set the threshold for rejecting the null hypothesis?
  – Choosing a lower threshold increases the chance of incorrectly accepting the null hypothesis.
  – Choosing a higher threshold increases the chance of incorrectly rejecting the null hypothesis.
  – A common compromise is to reject the null hypothesis if $p < 0.05$, but there is nothing magical about this number.

Hypothesis testing

- In most experiments we need more complex statistical analyses than the *t* test (e.g. ANOVA), but the logic is the same: Given certain assumptions, the test allows us to determine the probability that our results could have arisen by chance in the absence of the hypothesized effect (i.e. if the null hypothesis were true).

Fitting models

- Statistical analyses generally involve fitting a model to the experimental data.
- The model in a t-test is fairly trivial, e.g.

duration = μ + syllable_type (syllable_type is ‘open’ or ‘closed’)

Fitting models

- Statistical analyses generally involve fitting a model to the experimental data.
- The model in a t-test is fairly trivial, e.g.

  duration = μ + syllable_type (syllable_type is ‘open’ or ‘closed’)

  $\text{duration}_{ij} = \mu + \text{syllable\_type}_i + \mathit{error}_{ij}$

- Analysis of Variance (ANOVA) involves more complex models, e.g.

  $\text{dur}_{ijk} = \mu + \text{vowel}_i + \text{syll\_type}_j + \mathit{error}_{ijk}$

  $\text{dur}_{ijk} = \mu + \text{vowel}_i + \text{syll\_type}_j + \text{vowel*syll\_type}_{ij} + \mathit{error}_{ijk}$

- Model fitting involves finding values for the model parameters that yield the best fit between model and data (e.g. minimize the squared errors).
- Hypothesis testing generally involves testing whether some term or coefficient in the model is significantly different from zero.