24.963

Linguistic Phonetics

Basic statistics

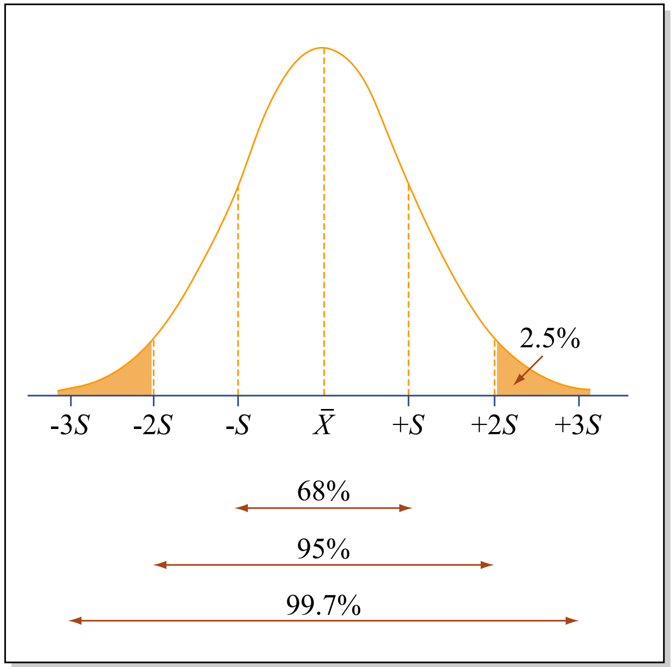


Image by MIT OCW.
Adapted from Kachigan, S. K. *Multivariate Statistical Analysis*. 2nd ed. New York, NY: Radius, 1991.

Assignments:

- Send me a paragraph on your final project (due 11/24)
- Write up the affricates experiment (due 12/1)

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

Measures of central tendency:

- Mean: $$M = \frac{\Sigma x_i}{N}$$
  - M is used for sample mean, μ for population mean.

- Median: The value that separates the lower half of a set of observations from the higher half.
  - Arrange the values from low to high. The median is in the middle of this list.

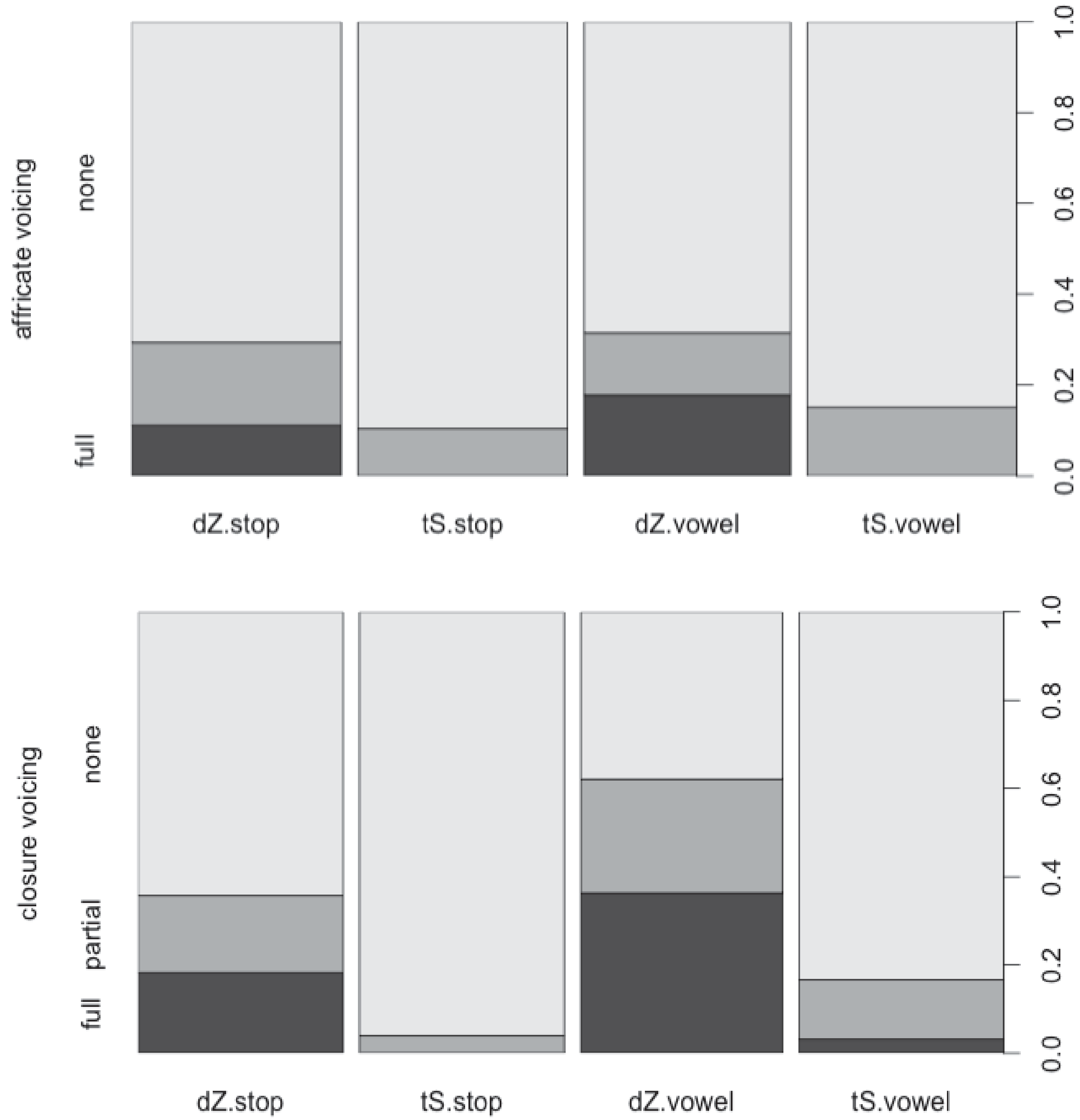
Descriptive statistics

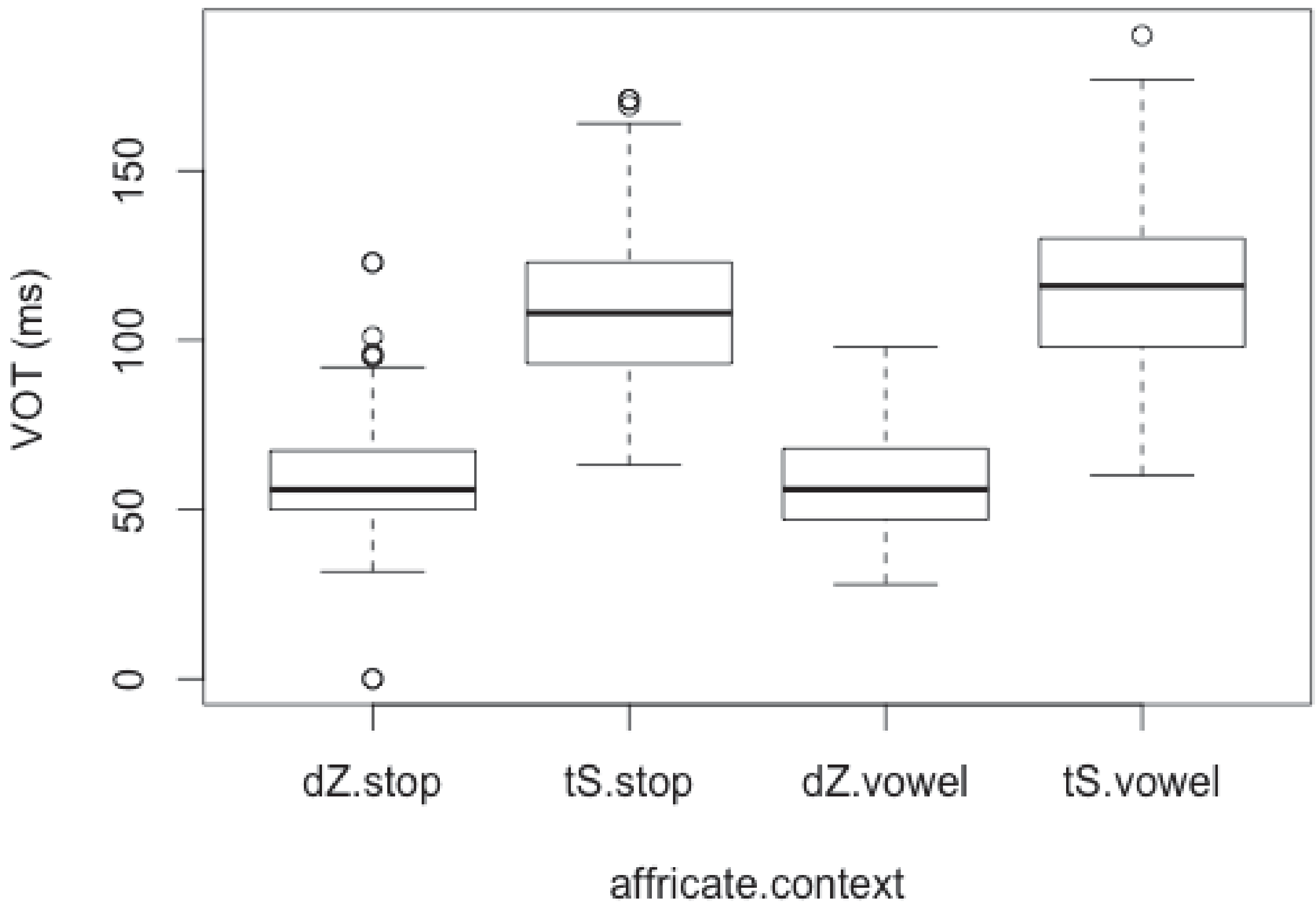
A measure of dispersion:

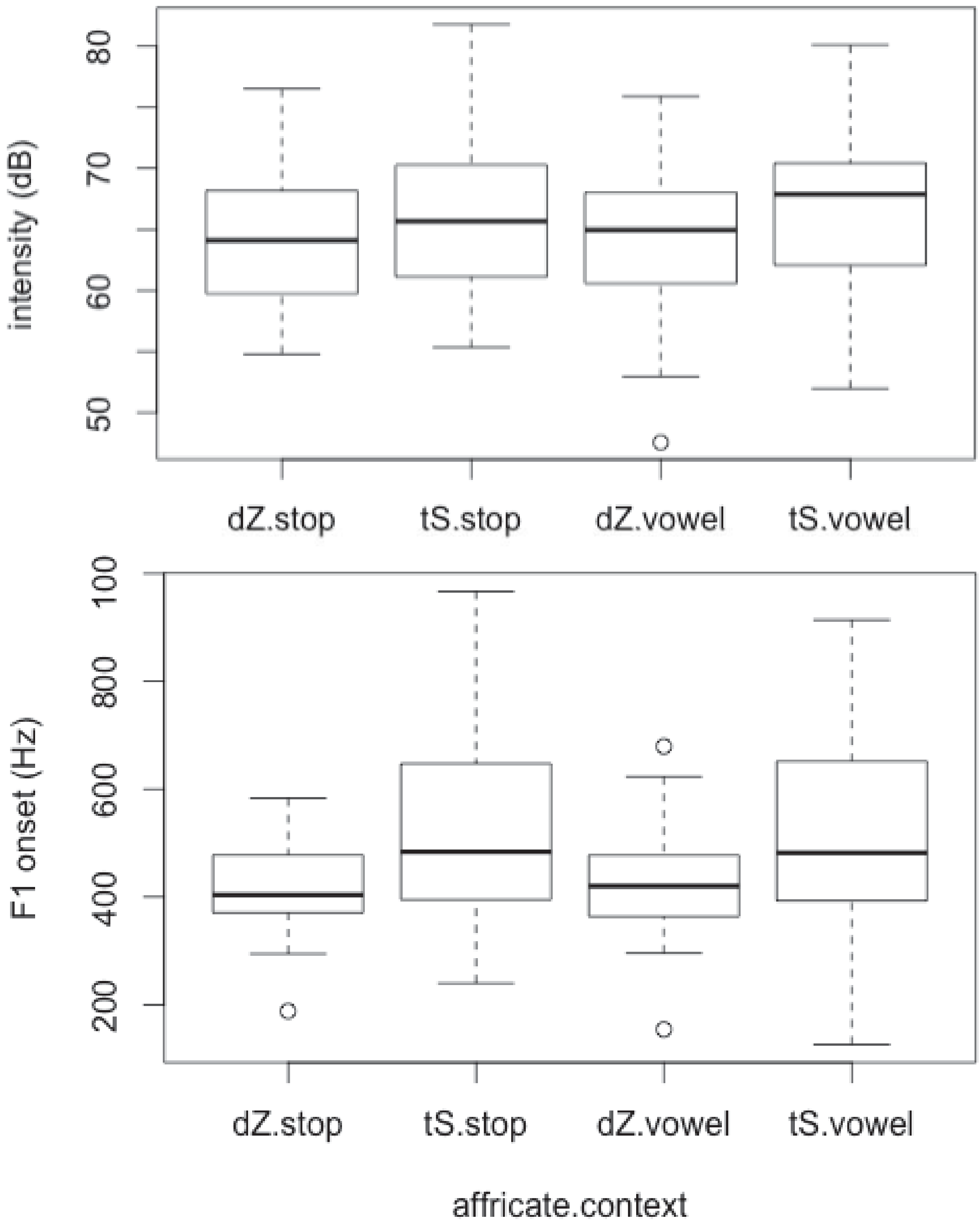
- Variance: mean of the squared deviations from the mean

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

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


affricate voicing
none
full
dZ.stop
tS.stop
dZ.vowel
tS.vowel
0.0
0.2
0.4
0.6
0.8
1.0
closure voicing
none
partial
full
dZ.stop
tS.stop
dZ.vowel
tS.vowel
0.0
0.2
0.4
0.6
0.8
1.0


VOT (ms)
0
50
100
150
dZ.stop
tS.stop
dZ.vowel
tS.vowel
affricate.context


intensity (dB)
80
70
60
50
dZ.stop
tS.stop
dZ.vowel
tS.vowel
F1 onset (Hz)
100
800
600
400
200
dZ.stop
tS.stop
dZ.vowel
tS.vowel
affricate.context

Hypothesis Testing

Which properties differentiate /tʃ/ and /dʒ/?

	mean VOT (ms)	mean intensity (dB)	mean F1 onset (Hz)
/tʃ/	112	67	531
/dʒ/	58	64	423

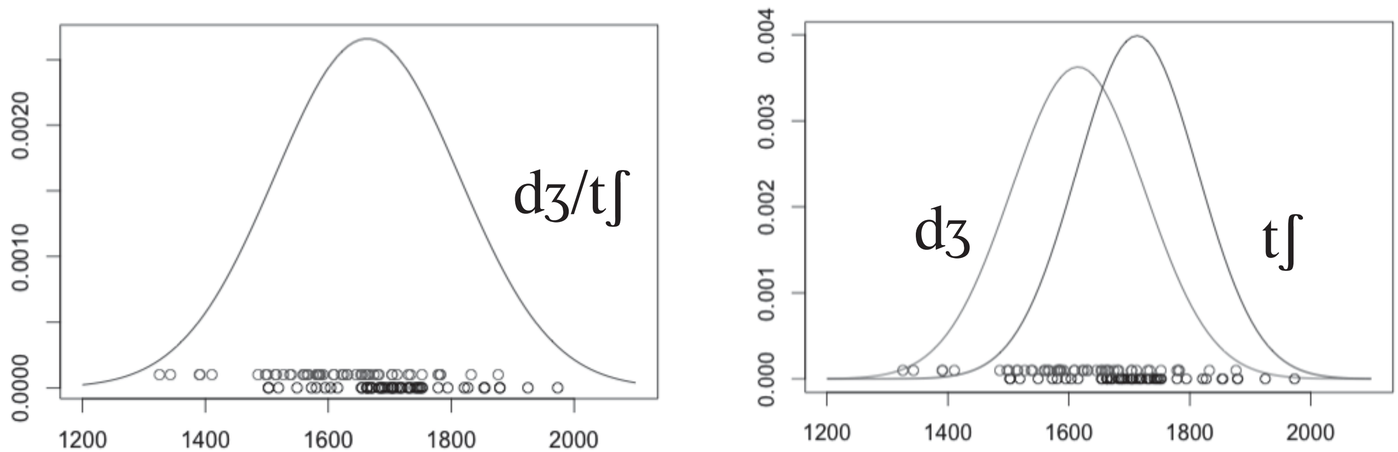
Hypothesis Testing

- Is peak intensity of frication in /tʃ/ higher than in /dʒ/? I.e. is 67 dB significantly higher than 64 dB?

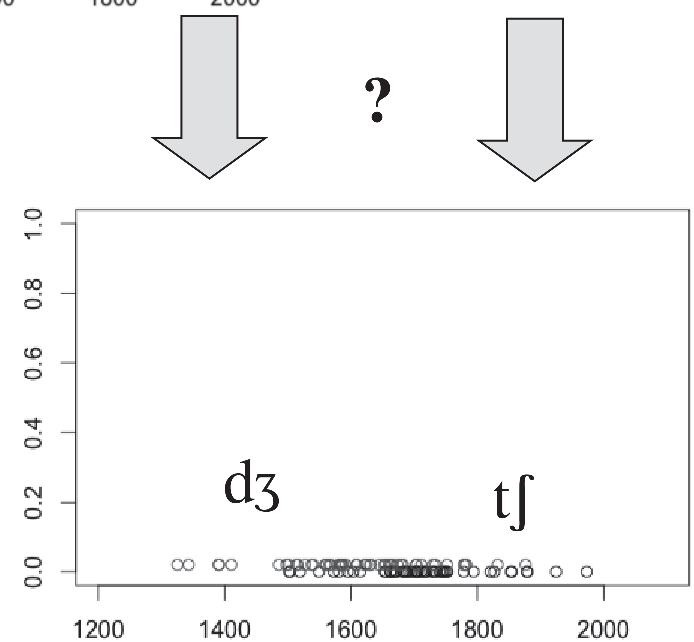
- Could the apparent differences have arisen by chance, although the true (population) means of frication intensities in the two affricates are the same?

- I.e. given that intensity of /tʃ/ and /dʒ/ frications varies, we might happen to sample most of our /tʃ/ tokens from the high end of the distribution, and most of our /dʒ/ tokens from the low end.
- Statistical tests allow us to assess the probability that this is the case.

- Could the apparent differences have arisen by chance, although the true (population) means of /tʃ/ and /dʒ/ are the same?



- 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.
  - The mean frication intensity of /tʃ/ differs from the mean frication intensity of /dʒ/ (Mean difference ≠ 0)

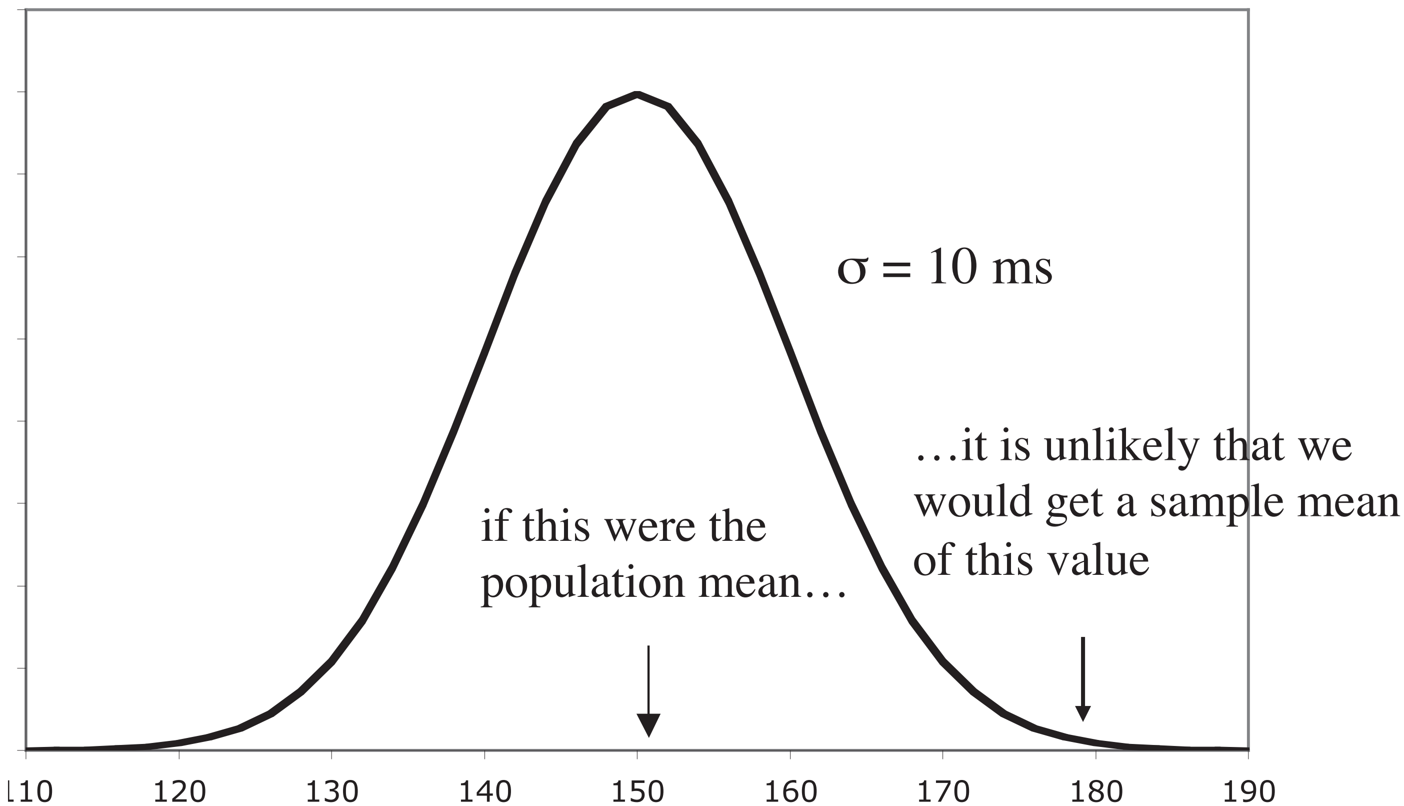
- We actually evaluate two exhaustive and mutually exclusive hypotheses, a **null hypothesis** that the mean has a particular value, and the alternative hypothesis that the mean does not have that value.
  - The mean frication intensity of /tʃ/ = the mean frication intensity of /dʒ/ (Mean difference = 0).
- 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.

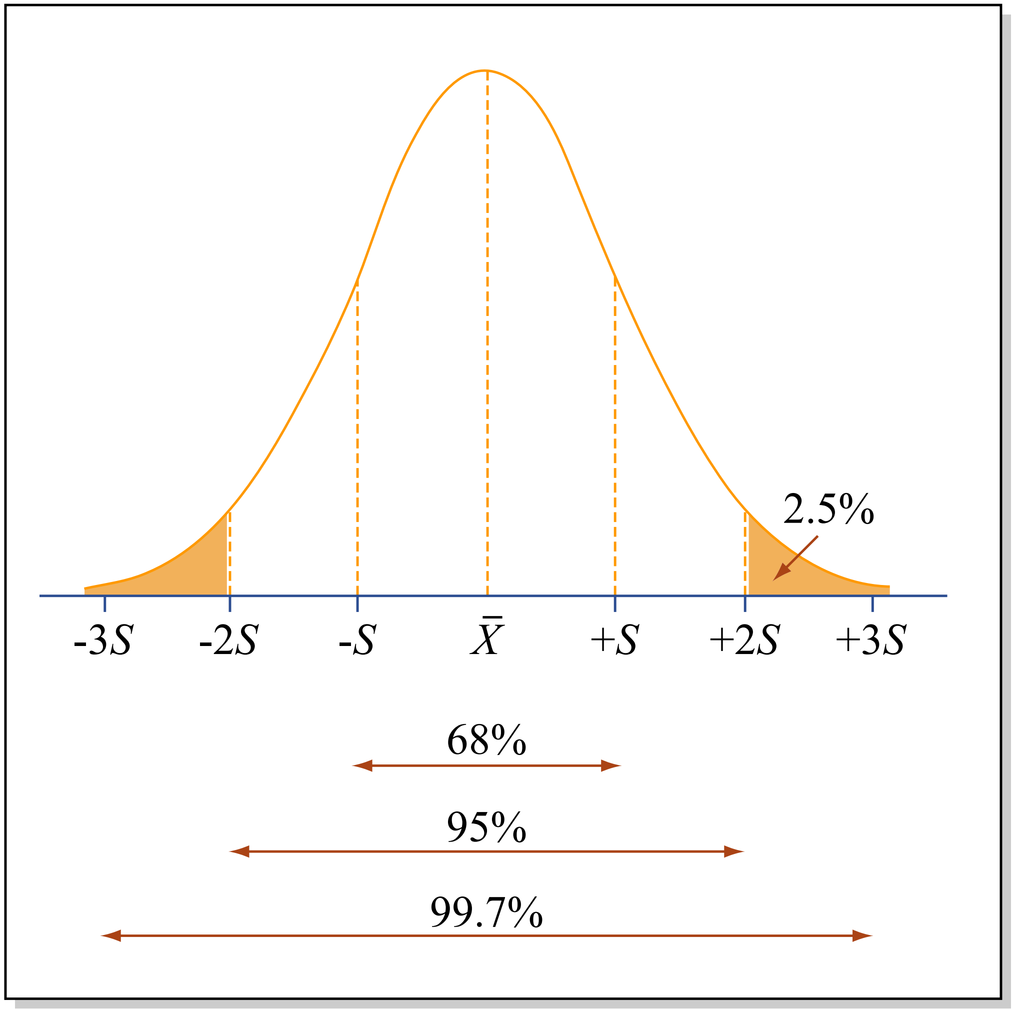


Image by MIT OCW.
Adapted from Kachigan, S. K. *Multivariate Statistical Analysis*. 2nd ed. New York, NY: Radius, 1991.

Distribution of Sample Means

- Basic assumption: The distribution of sample means is normal.
- This is guaranteed to be true if the population from which the sample is taken is normally distributed.
- But the distribution of sample means is approximately normal even if the population is non-normal, as long as the samples are large enough.

- http://onlinestatbook.com/stat_sim/sampling_dist/index.html

Hypothesis Testing

- 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:

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

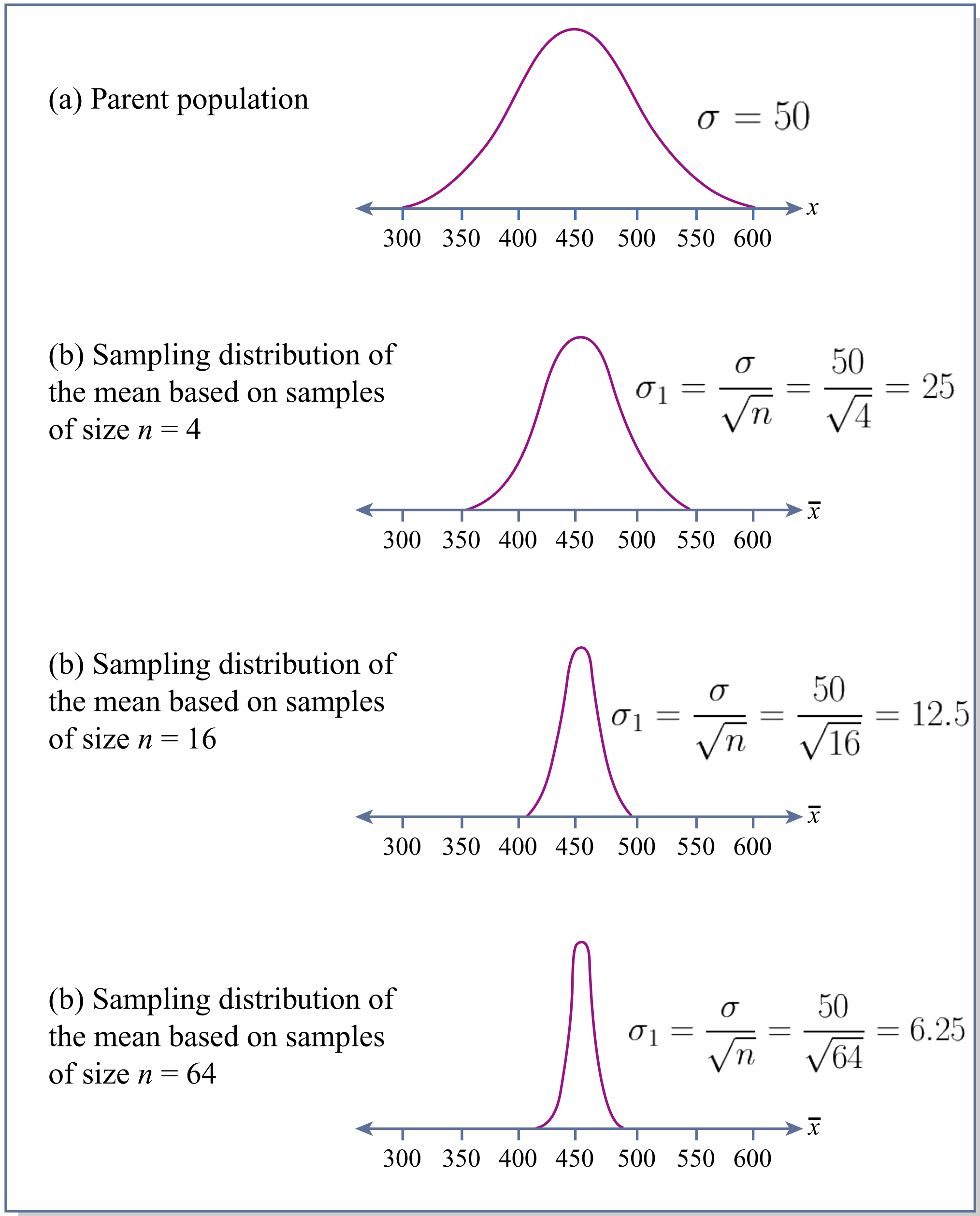


Image by MIT OCW.
Adapted from Kachigan, S. K. *Multivariate Statistical Analysis*. 2nd ed. New York, NY: Radius, 1991.

Hypothesis Testing

- The mean of the distribution is determined by hypothesis.
  - E.g. mean = 1500 Hz or mean difference = 0.
- Population variance is estimated from the sample variance. Unbiased estimate of the population variance:

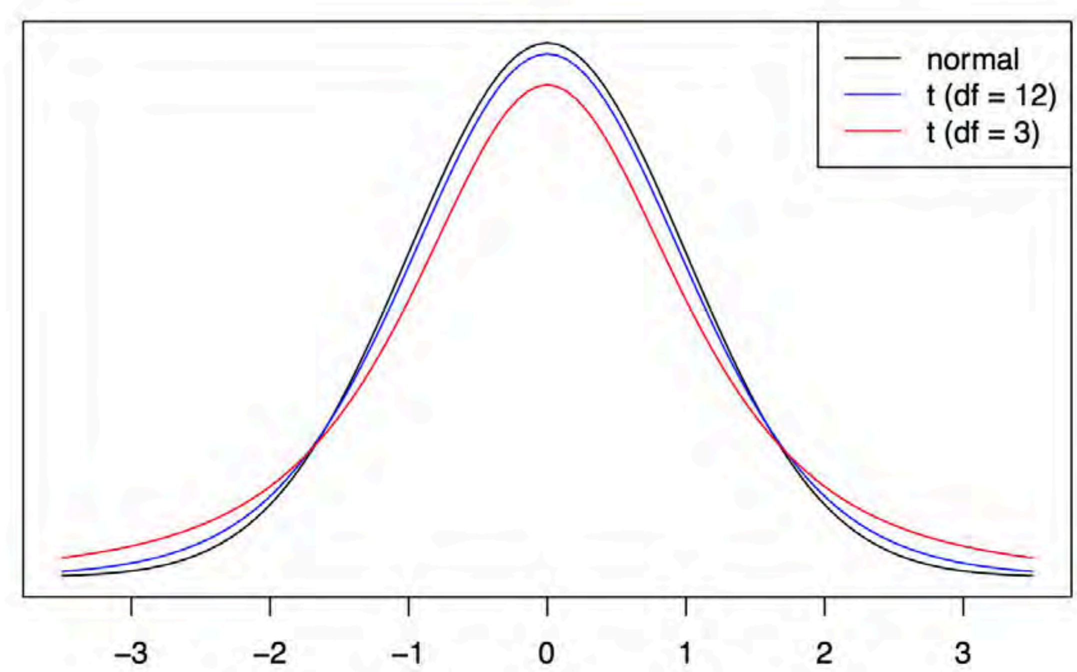
$$S^2 = \frac{\Sigma(x_i - M)^2}{N-1}$$

  - 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).



t test for independent means

- When we compare means, we are actually sampling a population of differences (e.g. differences in intensity of frication of /tʃ/ and /dʒ/).
- 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 and sizes of the two samples.
  - Or, if the observations are paired, based on the variance of the differences ('paired t-test')
- Using a paired t-test (ignoring repeated measures):
  - $S_M = 3.34/\sqrt{252} = 0.21$
  - $M = M_1 - M_2 = 66.7-64.3 = 2.4$
    (difference between sample means)
  - $\mu = 0$ (null hypothesis is that the difference between pop. means is 0).

$$t(22) = \frac{M-\mu}{S_M} = \frac{2.4-0}{0.21} = 11.4$$

$$p(\left|t(252)\right| \geq 11.4) = 2.56 \times 10^{-6}$$

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.

affricate VOT = μ + voice   (voice is ‘voiced’ or ‘voiceless’)

Fitting models

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

  affricate VOT = μ + voice (voice is 'voiced' or 'voiceless')

  $VOT_{ij} = \mu + voice_i + \mathit{error}_{ij}$

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

  $VOT_{ijk} = voice_i + context_j + \mathit{error}_{ijk}$

  (context takes different values for /stop_, /vowel_)

  $VOT_{ijk} = voice_i + context_j + voice{*}context_{ij} + \mathit{error}_{ijk}$

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