The p value

- Example: The mean change in blood pressure after treatment in 36 patients with hypertension is 5.0mmHg with standard deviation = 15mmHg. Such data may arise when the subject serves as his or her own control. (e.g., blood pressure may be recorded before commencement of treatment, and then after 1 week of treatment.)
- Null hypothesis: there is no effect of treatment on blood pressure.

Paired differences

- If the change in B.P. d is calculated for each patient and the null hypothesis is true, then the 36 d's should be close to zero.
- The d's are termed the paired differences.

  $\bar{d} = \Sigma d/n = 5.0 \quad SD = (\Sigma(d-\bar{d})^2/(n-1))^{½} = 15.0$
- Thus, $SE(\bar{d}) = SD(d)/sqrt(n) = 2.5$

  and $z = 5.0/2.5 = 2.0$
- z determines p as 0.046 (from normal dist'n table)
- A statistical significance test considers this p-value.

Statistical Inference

- Hypothesis Testing: method of deciding whether the data are consistent with the null hypothesis.
- Given a study with a single outcome measure and a statistical test, hypothesis testing can be summarized in three steps:
  1. Choose a significance level, α, of the test.
  2. Conduct the study, observe the outcome, and compute the p-value.
  3. …

Statistical Inference

3. If the p-value <= α → data are not consistent with the null hypothesis.

If p-value > α, do not reject the null hypothesis, and view it as "not yet disproven".

- Do not confuse the significance level and the p-value!
- If one rejects the null hypothesis when it is in fact true, then one makes a Type I error.
- The significance level α is the probability of making a Type I error. This is set before the test is carried out. The p-value is the result observed after the study is completed.

Small samples of continuous data

- Student's t-distribution: If samples are small, $\overline{x}$ and s will not be close to μ and σ.
- (In one way sample size is already taken into account through the calculation of the standard deviation of the mean, SE($\overline{x}$) when dividing by sqrt(n).)
- In small samples, however, values of s far from σ are not uncommon.

Small samples of continuous data

- → although $\overline{d}$ will still have a Normal distribution, it can no longer be assumed the ratio $\overline{d}/SD(\overline{d})$ will.
- Our previous analyses were based on the assumption that s (sample) was close to σ (population).
- We must modify the calculation of both the p-value and a confidence interval.

Small samples of continuous data

- First, relabel the ratio as **t** instead of **Z** to avoid confusion.
- The confidence interval Z_α is replaced by t_α, to yield

$$\overline{d} \pm t_\alpha \times SE(\overline{d})$$

- The ratio $\overline{d}/SE(\overline{d})$ is then known as the Student's t-statistic and under the null hypothesis is assumed to be distributed as the Student's t-distribution.

Small samples of continuous data

- t_α depends not only on α but also on the number of degrees of freedom, df, on which σ is estimated.
- The number of degrees of freedom depends on two factors:
  1. The number of groups we wish to compare;
  2. The number of parameters we need to estimate to calculate the standard deviation of the contrast of interest.
- (calculation of df depends on the particular problem…)

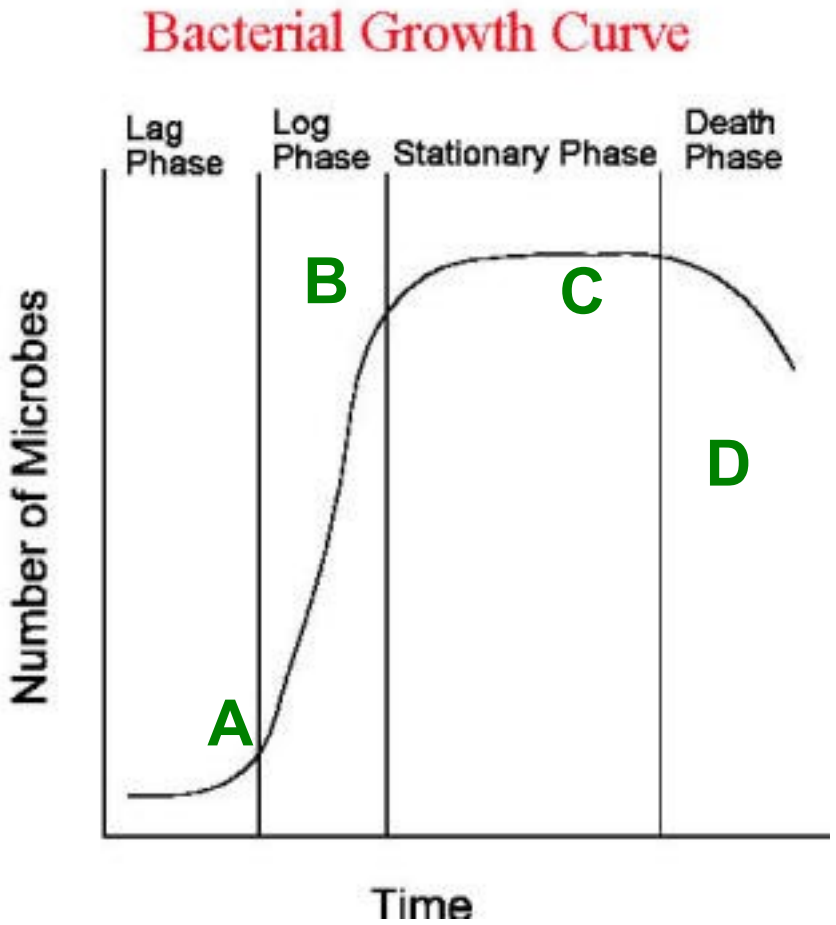
The t-statistic

- Given df and α, t_α is obtained from the Student's t-distribution table.
- For df = inf, t_α is equal to Z_α as read off of the Normal distribution table.
- t_α ↓ as df ↑, reflecting the increasing uncertainty concerning the estimate of s as sample sizes get smaller.
- There are two main types of t-tests: Paired and Unpaired.

Two types of clinical trial

- Cross-over trial: subjects receive both the test and control treatments in a randomized order. Useful in stable, chronic diseases such as diabetes or arthritis, where the purpose of treatment is palliation and not cure.
- Matched case-control study: identifies group of persons with the disease, and control group without the disease, and examines the relationship of a risk factor to the disease. Control group is picked to resemble diseased group w.r.t. certain risk factor (e.g., % of smokers).

Q1: If an adaptive step size integration is performed, which region of the function will permit the largest step size?



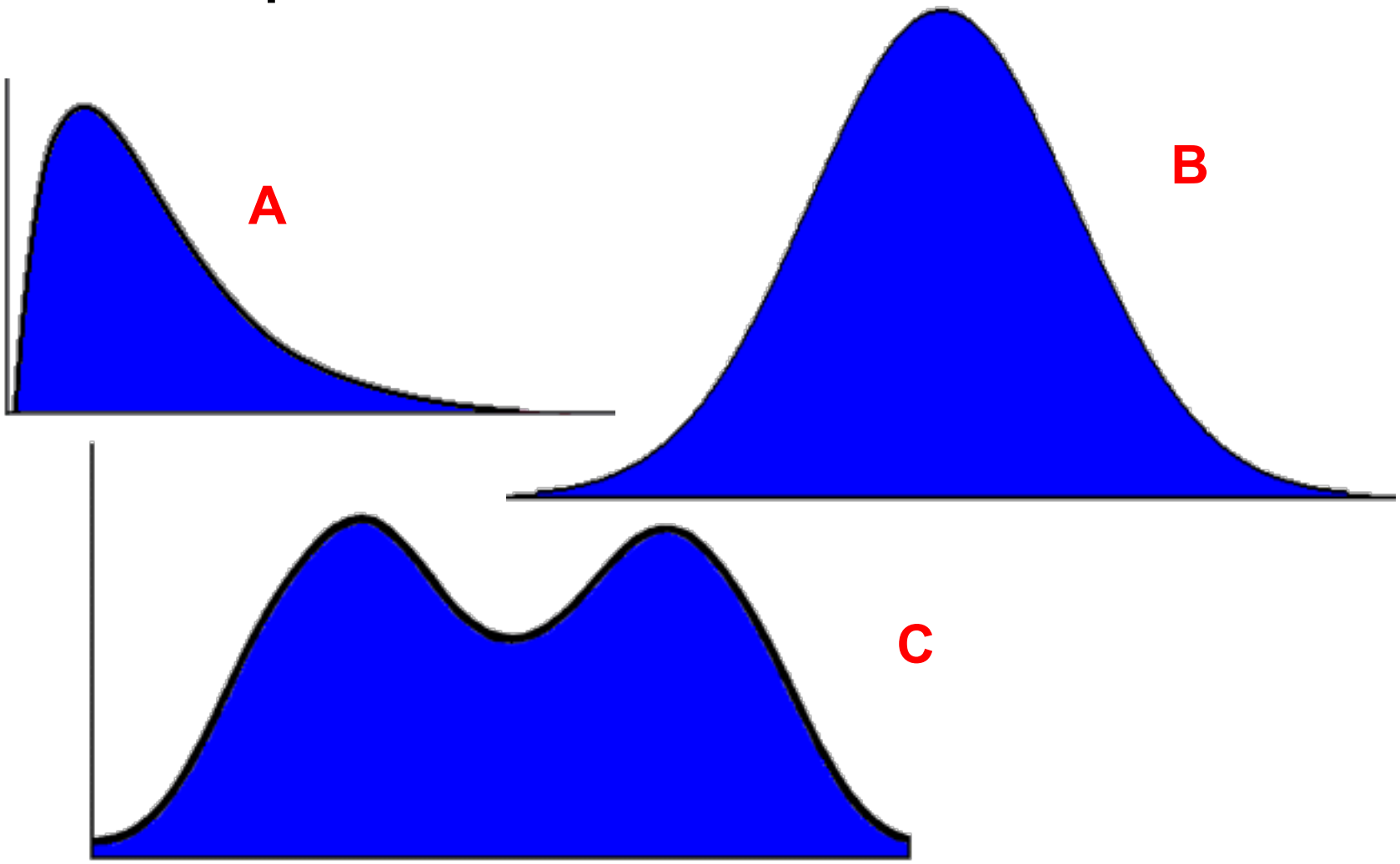
Q2: Which of the following is NOT a “null hypothesis”?

A. Diabetic patients do not have raised blood pressure.

B. Oral contraceptives do not cause breast cancer.

C. Smoking cigarettes increases the risk of heart disease.

Q3: Which of the following is an appropriate population for a two sample z-test?



Experiment

- Develop hypotheses
- Collect sample/Conduct experiment
- Calculate test statistic
- Compare test statistic with what is expected when H_0 is true
  - Reference distribution
  - Assumptions about distribution of outcome variable

Example: Hypertension/Cholesterol

- Mean cholesterol hypertensive men
- Mean cholesterol in male general population (20-74 years old)
- In the 20-74 year old male population the mean serum cholesterol is 211 mg/ml with a standard deviation of 46 mg/ml

Cholesterol Hypotheses

- H_0: $\mu_1 = \mu_2$
- H_0: $\mu = 211$ mg/ml
  - μ = population mean serum cholesterol for male hypertensives
  - Mean cholesterol for hypertensive men = mean for general male population
- H_A: $\mu_1 \neq \mu_2$
- H_A: $\mu \neq 211$ mg/ml

Cholesterol Sample Data

- 25 hypertensive men
- Mean serum cholesterol level is 220mg/ml

  ($\bar{X}$= 220 mg/ml)

  – Point estimate of the mean
- Sample standard deviation: s = 38.6 mg/ml

  – Point estimate of the variance = s^2

Experiment

- Develop hypotheses
- Collect sample/Conduct experiment
- Calculate test statistic
- Compare test statistic with what is expected when H_0 is true
  - Reference distribution
  - Assumptions about distribution of outcome variable

Basic Research Terminology

- **Retrospective:** (historic) look back at events that have already taken place
- **Prospective:** Watch for future outcomes
- **Case Control Study:** Persons w/ disease & those w/out are compared
- **Cohort Study:** Persons w/ and/or w/out disease are followed over time

Terminology (Cont.)

- **Cross-sectional Study:** Presence or absence of exposure to possible risk factor measured at one point in time. Prevalence obtained.
- **Prevalence:** The # of new cases and existing cases during specified time period.
- **Incidence:** The # of NEW cases per unit of a population at risk for disease occurring during stated time period.

Historical Minute
First "Clinical Trials"

Historical Highlights of Drug Trials

- 1909: Paul Ehrlich – Arsphenamine (syphilis)
- 1929: Alexander Fleming – Penicillin (1st bacterial)
- 1935: Gerhard Domagk – Sulfonamide (bacterial)
- 1944: Schatz/Bugie/Waksman – Streptomycin (TB)
- By 1950, the British Medical Res. Council developed a systematic methodology for studying & evaluating therapeutic interventions

Core Components of Clinical Trials

- Involve human subjects
- Move forward in time
- Most have a comparison CONTROL group
- Must have method to measure intervention
- Focus on unknowns: effect of medication
- Must be done before medication is part of standard of care
- Conducted early in the development of therapies

Core Components of Clinical Trials

- Must review existing scientific data & build on that knowledge
- Test a certain hypothesis
- Study protocol must be built on sound & ethical science
- Control for any potential biases
- Most study medications, procedures, and/or other interventions

Simplified

- **Randomized:** Schemes used to assign participant to one group
  - Ex: Every 3 gets higher dose
- **Nonrandomized:** All with Hep. C = cases; others = controls
- **Protocol:** Study design - instructions
- **Blinded:** Participants do not know if in experimental or control group
- **Double Blinded:** Participants AND staff do not know group assignment
- **Placebo:** Inactive pill w/ no therapeutic value

Components of Clinical Trial Protocols

- Investigating two or more conditions so have two(+) groups
  - Ex: drug vs. placebo; medicine vs. surgery; low dose vs. high dose
- Specific inclusion/exclusion criteria
- Sample size & power calculations
- Plan re: potential biases
- Plan re: handling of attrition/loss to follow up

Study Participant Recruitment

- Identify eligible participants
- Explain study
- Provide informed consent
- Reassess eligibility
- Assign to one group

Participants should be told:

- May have side effects (adverse effects)
- Time commitment
- Benefits & risks
- May withdraw at any time
- Enrollment 100% voluntary

Phases of Clinical Trials

- **Phase I:** Small group [20-80] for 1st time to evaluate safety, determine safe dosage range & identify SE
- **Phase II:** **Rx/tx given to larger group [100-300] to confirm effectiveness, monitor SE, & further evaluate safety**

Phases of Clinical Trials (cont.)

- **Phase III:** Rx/tx given to even larger group [1,000-3,000] to fulfill all of Phase II objectives & compare it to other commonly used txs & collect data that will allow it to be used safely
- **Phase IV:** Done after rx/tx has been marketed - studies continue to test rx/tx to collect data about effects in various populations & SE from long term use.

Summary of Phases I-III

	# Subs.	Length	Purpose	% Drugs Successfully Tested
Phase I	20 – 100	Several months	Mainly Safety	70%
Phase II	Up to several 100	Several months-2 yrs.	Short term safety; mainly effectiveness	33%
Phase III	100s – several 1000	1-4 yrs.	Safety, dosage & effectiveness	25-30%

Issues in Clinical Trials: Use of Placebo Trials

On international realm, 1999 "Declaration of Helsinki" revised to address use of placebos:

- Placebos not ethical in virtually all studies that involve diseases with PROVEN tx
- Remain ethical in trials where no proven tx
- Revisions due to controversy over use of placebos in attempting to find easy/cheap way to reduce HIV perinatal transmission
    - 1998 study in Ivory Coast, Uganda, & Thailand: HIV+ pregnant women given either placebo or shorter course of AZT

t-test example

- Hindmarsh and Brook (1987) give the heights of 16 children before treatment and one year after treatment with a growth hormone.
- Results were standardized for age (observed height minus predicted height for age divided by SD of height).

t-test example

Subject	Baseline	At 1 year	Difference
1	-0.7	4.1	4.8
2	0.0	3.4	3.4
3	-0.7	3.1	3.8
4	-0.7	3.0	3.7
5	0.5	2.8	2.3
6	-0.7	2.7	3.4
7	-0.6	2.5	3.1
8	-0.3	2.3	2.6
9	-0.7	2.2	2.9
10	-0.7	2.0	2.7
11	-0.5	1.8	2.3
12	-0.7	1.6	2.3
13	-0.5	1.3	1.8
14	-0.7	0.9	1.6
15	-0.4	0.8	1.2
16	-0.3	0.3	0.6
Mean:	-0.48	2.18	2.66
Median:	-0.65	2.25	2.65
SD:	0.33	1.03	1.06

t-test example

- The distribution is skewed, as demonstrated by the coefficient

$$sk = 3(\text{mean} - \text{median})/SD$$

$$= 1.55$$

- Which suggests that the data are not normally distributed.
- Note that just because the basic observations appear not to have normal distributions, this does not necessarily mean that our methods do not apply!

t-test example

- In this case it is the differences before-after, that have to be checked.
- In fact the differences appear to be symmetric (sk = 0.03).
- SD(d) = 1.06, and SE($\overline{d}$) = 1.06/sqrt(16) = 0.265
- This is a small study so we need to use the confidence interval with t_{α}, not Z_{α}.
- In this example the data are paired and the degrees of freedom are **df = n – 1** = 15

t-test example

- 95% confidence interval for the mean difference is:

$$2.66 \pm (t_{0.05} \times 0.265)$$

- From Student's t-distribution table, with df=15, $t_{0.05} = 2.131$.
- → 2.10 – 3.22 cm.
- We can calculate the Student's t-statistic as t=2.66/0.265=10.0.

t-test example

- From the table, the largest value with 15 degrees of freedom is 4.073, corresponding to α=0.001.
- Thus, $p<0.001$.
- We conclude that the growth hormone was effective in increasing the stature of these children.