Today: Two hard ideas

1. Experimental strategy using randomization
2. Judging when apparent differences are "real"

both illustrated through the Food Preference Study (pp 161-169).
Also, Assignment 2 will be displayed at the end of the lecture (and is already posted on the web.)

Recall from the article on “Measuring the Effects of Social Innovations by Means of Time Series”, a preliminary discussion of “designed experiments”.
Experimental group (E) vs Control Group (C)
Randomization of subjects to E or C.
Guarantees the statistical balance of the comparison groups before “treatment”. This design is necessary to infer causation …

Example from Criminology:

Pre-trial hearing – does it save the public expense on the trial process?
Series of cases. Coin toss to allocate to pre-trial hearing (E) or not (C).
Cost per case in each group subsequently calculated.
Coin toss achieves randomization. Is it necessary?
Other methods of group formation:
April (C) , May (E) : but follow-up behavior might be different.
Firm 1 (C) Firm 2 (E): but cases might be different.
Last name A-M (C), N-Z(E): ethnic bias?
Alternating C and E: often works if order “random”.
Randomization is best generally applicable method for unambiguous results.

Food Preference Study (pp 161-169) (which you have read)

Nutrition test with three diets: control, diet 1 and diet 2

Blocking strategy: ensure equivalence at start

Randomization strategy: even out unintended treatment differences

Fig 1. Obvious that H is better than Casein.

Why a straight line fit? P 165 Numerical summary in one number.

3.72 for Liquid H

3.66 for Solid H

2.91 for Casein (cheese protein)

Interpretation: H better because growth rate higher per gm intake

H also apparently more palatable (to rats!).

Palatability Study (for people).

Ordered rating – see pics p 166

Male, Female frequency distribution of ratings (Table 1 p 166)

Chance differences vs reproducible differences

Jargon: Statistical Inference – deciding when apparent differences are large enough to suggest they are not merely chance.

"If something happens that is unusual under ordinary circumstances, then this is evidence that the circumstances are not ordinary."

Here: Ordinary circumstances = diets make no difference to preference

Unusual = something that happens less than once in 20.

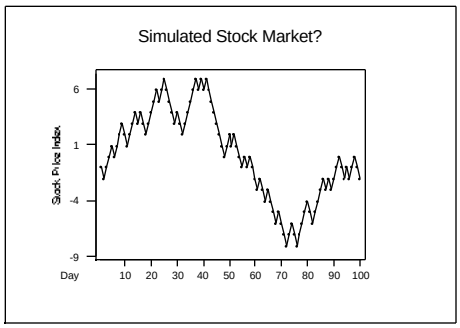
Something =

averages of preference scores are C=.22, LH=1.18, SH=1.06

The assessment of "unusual" (difference at least .45) depends on the SD of the ratings within diets. How this is determined is not explained at this point.

Assignment 2 due Wed Sept 25 (4:30 pm in boxes outside of K 9509)

1. What is unrealistic about the following simulation of the stock market index, and how can it be made more realistic?



2. Explain how a simulation of the stock market index might help us to avoid irrational predictions of the future trend of the index.

3. Suppose you have the job of examining the past five years of monthly automobile sales in BC. You are asked to determine those months that had higher than expected sales and those that had lower than expected sales. Your boss wants this information to compare with other economic indicators that were available over the same period - the boss's aim is to try to infer possible links between the economic series and the auto sales series. Is there a role for seasonal adjustment of the series for your analysis of car sales? Explain why or why not. (Assume that there IS a seasonal trend to auto sales data.)

4. The birth and death time series for 29 countries gave birth and death rates for 1975, 1980, 1985, 1990, 2001 and 2002. You were asked to think about how you might summarize this data and there was a suggestion of computing the ratio of the 2002 figure to the 1975 figure. (The data may be viewed at http://www.infoplease.com/ipa/A0004395.html although you are not expected to do any analysis of it for this exercise). What would you do with the ratio data to construct a useful graphic for studying this data set? (Assume your interest is in studying population changes by country).
Note about assignment returns: Pick up from far corner of statistics workshop room K9516 (inside door K9510). Mark is on inside of assignment. If you need privacy, please give Lab Instructor Robin Insley self-addressed envelopes. (Leave at STAT general office K 10545).
Turn around time usually 1 week or less.

Next topic (see Course Outline):"Risk" - more simulations.

Appendix: Equation of a straight line

Think of Y measured on vertical axis, X on horizontal axis.

$Y = a + m X$

a = value of Y when X=0 "Intercept"
m = rise in Y per unit rise in X "Slope"
Line below is Y=5+2X