Beta

Administrative Issues

Feedback exam time and location

– Sunday afternoon, 3 p.m.

– Location determined by cohort (see Stat 603 web page)

– About 35 questions, multiple choice

Practice

– Last year's exam on web page.

– Hard part is figuring out which method is best

Summary of Key Take-Away Points

Normality

– Why it is useful (empirical rule)

– Why it is common (central limit theorem)

– When it sometimes fails as a model (extreme extrapolation)
Quantile plot as a diagnostic tool.

Standard error

– As a measure of the variability of means (control charts)

– Understanding the factors that affect amount of variation.
(role in design of studies, like margin for error in survey)

– Magical perspective: estimating sample-to-sample variation from information in only one sample.

Optimal decisions

– Role of microeconomics and elasticity.

– Trading cost of one error for another.

– Demands for information

Confidence interval

– Combines normality (via CLT) and standard error.

– Canonical form estimate ± 2 standard errors of the estimate.

– Values that are outside the CI (like zero) are not plausible.

– Examples for means, proportions and differences of these.

Hypothesis testing

– Relationship to confidence intervals.

– Counting standard errors: t-ratio or t-stat = # SE from zero.

– p-value

– Errors that occur when using statistical procedures
Type I, Type II and type 3 errors

– Role of assumptions (independence, equal variance, normality)
How to check for these.

Surveys

– Basic ideas of random surveys.

– Questions to ask of any survey (e.g., non-response, question wording)

– Subtle issues (length-biased sampling)

Some fundamental tests (as examples)

– How to compare means of two samples.

– Effects of doing many comparisons.

Covariance and correlation

– As measures of dependence

– Role in formulation and understanding returns, risk of portfolio.

Key Application for Today

How much of a stock do I want to buy?

– Do stocks really do that well over time?

– What stocks make a good portfolio? "not all eggs in one basket"

– How well do the "best" portfolios perform?

Definitions and Concepts

Regression to the mean

– In the presence of random variation, the subject with the best score declines toward the mean when measured a second time.

– Impact on management practice.
Deming lectures on the importance of separating random from systematic variation.

Return and risk

– Return = rate of growth, "interest" on assets (average return)

– Risk = variability, chance for less tomorrow than today.

– Example: What is the effect of a sequence of alternating 10% increases and 10% decreases on an investment over time?

Covariance and correlation

– Related measures of association designed for normality

– Role in the design and use of portfolios
A portfolio is invests a given quantity of money in a collection of stocks (e.g.) with varying amounts of money in the constituent stocks.

– Covariance "algebra": variance of sums depends on covariance

– Correlation removes dependence of covariance on scale:

$$-1 \quad \text{Correlation} \quad +1$$

"What's a big covariance?" is a hard question to answer. It is easy to recognize a big correlation.

– Size of correlation related to accuracy of prediction in the method know as "regression" that is covered in Stat 621

Discussion

Uncorrelated independent

– Correlation only measures linear association

Covariance measures linear association

– Task in Stat 621 is to "explain" variation

– Covariance is a measure of how well it's been done
BUT only in the linear sense.

– Freedman text has a very special introduction to these ideas.

What's a big correlation?

– It all depends!

– Statistical significance (i.e., different from zero) matters, but

econometrics	0.9 common
social science	0.3 exciting

Examples

Performance of mutual funds (supplemental, mutfunds.jmp, p 216)

Can you pay attention to any of the claims of past performance?

– Why do year-to-year differences appear significant, but then change in sign? Useful example for the final assignment.

– Readings from Freedman about "regression to the mean."

Stocks, portfolios and the efficient frontier
(p 188, finmark.jmp, stockret.jmp, frontier.jmp, portfol.xls)

How can we choose a portfolio of stocks that maximizes our returns (earn the most interest) for a given level of risk?

Historical patterns...

– Are T-Bills really "risk-free"? (p 191)
– Are stocks really more "risky" than T-Bills? (p 193, 195)

Building small portfolios...

– How does one reduce risk by constructing a portfolio?
– Which items work best together in a portfolio? Combining
  - Sears and Penney, two retailers, does not reduce risk much.
  - Exxon and Amoco, both oil, does not reduce risk much.
  - Mixing JPMorgan and GM does. (p 200)

Why?

– Covariance and correlation, as seen in plots of returns.
– Algebra of covariances (p 204)
– Variance/covariance rules

(a) $\text{Var}(a X) = a^2 \text{Var}(X)$
$\text{Var}(-X) = \text{Var}(X)$

(b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ if independent
$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y)$

(a+b) $\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab\,\text{Cov}(X,Y)$

– Worked example: reproduce the properties of the mixed portfolio from covariances rather than from data directly.

Efficient frontier...

– What's the optimal weighting for a stock in a portfolio?
– Random portfolios reveal a barrier to performance (p 208)
– How does one achieve such a portfolio?

– Linear programming (i.e., the Excel solver) can be used to find the optimal weights.