2025 Level 1 - Quantitative Methods

Prerequisite Readings	**Page**
Interest Rates, Present Value, and Future Value	2
Organizing, Visualizing, and Describing Data	11
Probability Concepts	26
Common Probability Distributions	33
Sampling and Estimation	42
Basics of Hypothesis Testing	47

Interest Rates, Present Value, and Future Value

a. interpret interest rates as required rates of return, discount rates, or opportunity costs

b. explain an interest rate as the sum of a real risk-free rate and premiums that compensate investors for bearing distinct types of risk

c. calculate and interpret the future value (FV) and present value (PV) of a single sum of money, an ordinary annuity, an annuity due, a perpetuity (PV only), and a series of unequal cash flows

d. demonstrate the use of a time line in modeling and solving time value of money problems

e. calculate the solution for time value of money problems with different frequencies of compounding

f. calculate and interpret the effective annual rate, given the stated annual interest rate and the frequency of compounding

Time Value of Money

Page 1
LOS a
- interpret

- 3 rules of money: 1. money sooner is worth more than money later
2. larger cash flows are worth more than smaller cash flows
3. less risky cash flows are worth more than risky cash flows

- Interest rates (r) - can be thought of in 3 ways:

1/ required rate of return → the rate of return required by an investor or lender → moneytoday (1 + r) = moneytomorrow

2/ discount rate → the rate at which some future value is discounted to arrive at a value today → $\frac{\text{moneytomorrow}}{(1+r)} = \text{moneytoday}$

3/ opportunity cost → the value an investor or lender forgoes by choosing a particular action
i.e. r is the opportunity cost of current consumption

- typically: req. rate of return = discount rate = opportunity cost

Page 2
LOS b
- explain

→ suppose I lend you $1000 for one year. I will want:

r_f → real risk-free rate: single period rate
+ inflation premium → compensates for expected inflation (π^e)

$r_f + \pi^e$ = nominal risk-free rate
$[(1+r_f)(1+\pi^e)] - 1$

+ Default risk premium - compensates for credit risk

+ Liquidity premium → risk of loss vs. fair value if an investment needs to be converted to cash quickly

+ Maturity premium → greater interest rate risk (i.e. price risk) with longer maturities
→ will also have a premium for inflation uncertainty → the longer the time period, the more uncertain we are about the level of expected inflation

Page 3
LOS c
- calculate
- interpret

→ Future Value of a Single Cash Flow/

PV ——————— FV
t = 0 ——————— t = N

PV = present value
FV = future value
r = interest rate
N = # of periods
} r must be in the same periodicity as N

t = 0 —— r = 5% —— t = 1
100 (PV) —— FV = 100(1.05) = 105

e.g./ \$100, 5yrs., 6% → $100(1.06)^5$
semi-annual → $100(1.03)^{10}$
quarterly → $(1.015)^{20}$

t = 0 —— t = 1 —— t = 2
100 —— 100(1.05) —— $\underbrace{100(1.05)(1.05) = 100(1.05)^2}_{FV = PV(1+r)^N} = 110.25$

- the power of compounding/

25 yrs. ——————— 65yrs.
\$1000 ——————— FV = ?

S1
at 5%, N = 40: FV = $1000(1.05)^{40}$ = 7,039.99. (2000)
at 7%, N = 40: FV = $1000(1.07)^{40}$ = 14,974.46 (2800)
at 9%, N = 40: FV = $1000(1.09)^{40}$ = 31,409.42. (3600)

Page 4
LOS c
- calculate
- interpret

e.g./ \$5M at t = 0, r = 7% compounded annually, N = 5 years

Find FV: method 1: $5M(1.07)^5$ = 7,012,758.65
method 2: N = 5, I/Y = 7, PV = -5,000,000, PMT = 0
CPT FV = 7,012,758.65

e.g. 2/ Invest ¥2.5M, r = 8% compounded annually, N = 6 years
- Find FV: method 1: $2.5M(108)^6$ = ¥3,967,186
  method 2: N = 6, I/Y = 8, PMT = 0, PV = -2,500,000
  CPT FV = 3,967,186

e.g. 3/ \$10M at t = 5, E(r) = 9%, FV at t = 15?

0 —— 5 —— r = 9 —— 15
$\underbrace{5 \text{ — } 15}_{N = 10}$
10 M

method 1: FV = $10M(1.09)^{10}$ = 23,673,636.75
method 2: N = 10, I/Y = 9, PMT = 0
PV = -10,000,000
CPT FV = 23,673,636.75

$$PV_0 = \frac{10,000,000}{(1.09)^5} = 6,499,313.86$$

Page 5
LOS e
- solve

→ Frequency of Compounding/

- all rates are quoted annually → r_s (stated interest rate)

e.g./ PV = 10,000, N = 2, r_s = 8% compounded quarterly, FV = ?

$$FV = PV(1 + r_s/_m)^{mN} = 10,000(1 + .08/_4)^{4 \times 2} = 10,000(1.02)^8 = 11,716.59$$

or/

N = 2 × 4 = 8, I/Y = $^8/_4$ = 2, PMT = 0, PV = -10,000, CPT FV = 11,716.59

e.g. 2/ PV = $1M, N = 1, r = 6% compounded monthly, FV = ?

$$FV = 1M(1 + .06/_{12})^{12 \times 1} = 1M(1.005)^{12} = 1,061,678.81$$

or/ N = 1 × 12 = 12, I/6 = $^6/_{12}$ = .5, PMT = 0, PV = -1,000,000

CPT FV = 1,061,678.81

→ Continuous Compounding/ $FV = PV \times e^{rs \times N}$

e.g./ PV = 10,000, N = 2, r = 8% compounded continuously, FV = ?

$$FV = 10,000e^{.08 \times 2} = 11,735.11$$

or/ .08 × 2 = 2nd LN × 10 000 = 11,735.11

Page 6
LOS f
- calculate
- interpret

→ EAR: effective annual rate

$100 at 8%			EAR
	annual	100(1.08) = 108	8%
	semi-annual	$100(1.04)^2$ = 108.16	8.16%
	quarterly	$100(1.02)^4$ = 108.2432	8.2432%
	monthly	$100(1.00\dot{6})^{12}$ = 108.30	8.3%
	daily	$100(1.000219)^{365}$ = 108.3278	8.3278%
	continuous	$100e^{.08}$ = 108.3287	8.3287%

$[(1 + r/_m)^{mN} - 1]$ or $(e^{rN} - 1)$

- if we know EAR, we can solve for r_s

e.g./ EAR of 8.3%, compounded monthly

$$.083 = (1 + {}^{r_s}/_{12})^{12} - 1$$

$$1.083 = (1 + {}^{r_s}/_{12})^{12}$$

$$(1.083)^{1/12} = 1 + {}^{r_s}/_{12}$$

$$(1.083)^{1/12} - 1 = {}^{r_s}/_{12}$$

$$12\left[(1.083)^{1/12} - 1\right] = r_s$$ → 1.083 y^x (1 ÷ 12) - 1 × 12 =

$r_s = 8\%$

e.g. 2/ .083287 = $e^{r_s} - 1$

1.083287 = e^{r_s}

LN (1.083287) = r_s → r_s = 8%

Page 7
LOS c
- calculate
- interpret

→ Future value of a series of cash flows:

annuity - a finite set of level sequential cash flows

↳ ordinary → first cash flow at t = 1

↳ due → first cash flow at t = 0

→ Ordinary Annuity/

r = 5%

t = 0 | 1 | 2 | 3 | 4 | 5

-1000 -1000 -1000 -1000 -1000

$1000(1.05)^4$
$1000(1.05)^3$
$1000(1.05)^2$
$1000(1.05)^1$
1000
5,525.63

N = 5
PMT = -1000
I/Y = 5
PV = 0
CPT FV

$$FV = A\left[\frac{(1+r)^N - 1}{r}\right]$$

$$= 1000\left[\frac{(1.05)^5 - 1}{.05}\right] = 5,525.63$$

e.g./ $\underbrace{\text{€20,000/yr.}}_{\text{end of yr. CF}}$, N = 30, r = 9%

- ordinary annuity/

$$= 20,000\left[\frac{(1.09)^{30} - 1}{.09}\right] = 2,726,150.77$$

or/ N = 30, PMT = -20,000
PV = 0, I/Y = 9
CPT FV

Page 8
LOS c
- calculate
- interpret

→ Future value of a series of cash flows:

• Annuity Due/

r = 5%

t = 0 | 1 | 2 | 3 | 4 | 5

-1000 -1000 -1000 -1000 -1000

$1000(1.05)^5$
$1000(1.05)^4$
$1000(1.05)^3$
$1000(1.05)^2$
$1000(1.05)^1$
5,801.91

= ordinary annuity × (1.05)

• calculator in END mode
N = 5
PMT = -1000
PV = 0
I/Y = 5

$$\underbrace{\underbrace{FV = A\left[\frac{(1+r)^N - 1}{r}\right]}_{\text{ordinary ann.}} \times (1+r)}_{\text{annuity due}}$$

calculator in BGN mode
2^{nd} PMT 2^{nd} ENTER CE/C

Display END BGN.

N = 5, PMT = -1000, PV = 0, I/Y = 5
CPT FV

Page 9
LOS c
- calculate
- interpret

→ Future value of a series of cash flows:

- Unequal cash flows

r = 5%

t = 0	1	2	3	4	5
	-1000	-2000	-4000	-5000	-6000
	×	×	×	×	×
	$(1.05)^4$	$(1.05)^3$	$(1.05)^2$	$(1.05)^1$	$(1.05)^0$ = 19,190.76

→ Present value of a single cash flow:

r = 8%

t = 0 .... t = 6

PV = ? ← 100,000

$$PV = {100,000}/{(1.08)^6} = 63,019.96$$

(N = 6, PMT = 0, I/Y = 8, FV = 100,000)
CPT PV

$FV = PV\ (1 + r)^N$

$$PV = \frac{FV}{(1+r)^N}$$

$= FV\ (1 + r)^{-N}$

e.g./

r = 8%

t = 0 ... 4 ... 8

10 ... 100,000

$$PV_4 = {100,000}/{(1.08)^6}$$

$$= 63,016.96$$

$$PV_0 = {PV_4}/{(1.08)^4}$$

$$= 46,319.35$$

Page 10
LOS c
- calculate
- interpret

→ Present value of a series of cash flows:

- Ordinary annuity/

t = 0	1	2	3	4	5
	1000	1000	1000	1000	1000

r = 12%

$$PV = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right] \rightarrow 1000\left[\frac{1-\frac{1}{(1.12)^5}}{.12}\right] = 3,604.78$$

or/ N = 5
FV = 0
I/Y = 12
PMT = 1000 CPT PV

- Annuity Due

t = 0	1	2	3	4	5
1000	1000	1000	1000	1000	

$$1000 + 1000\left[\frac{1-\frac{1}{(1.12)^4}}{.12}\right] = 4037.35$$

or/ N = 4
PMT = 1000
I/Y = 12
FV = 0
CPT PV
+1000

e.g./

t = 0	1	2	...	19	20
200k	200k	200k		200k	

R = 7%

$2M today or 20 PMTs?

N = 19, PMT = 200,000, I/Y = 7, FV = 0
CPT PV (+ 20,000) = 2,267,119.05

or/

$$200k + 200k\left[\frac{1-\frac{1}{(1.07)^{19}}}{.07}\right] = 2,267,119.05$$

→ Present value of a series of cash flows:

Page 11
LOS c
- calculate
- interpret

e.g./

1M 1M 1M (30 PMTs)

t = 0 t = 9 10 11 12 t = 39

PV_0 = ?

r = 5%

PV_{10} → annuity due

PV_9 → ordinary annuity

$$PV_{10} = 1M\left[\frac{1-\frac{1}{(1.05)^{29}}}{.05}\right] + 1M$$

$$= 16,141,073.58$$

$$PV_0 = {PV_{10}}/{(1.05)^{10}} = 9,909,219$$

$$PV_9 = 1M\left[\frac{1-\frac{1}{(1.05)^{30}}}{.05}\right] = 15,372,451.03$$

$$PV_0 = {PV_9}/{(1.05)^9} = 9,909,219$$

BGN mode

N = 30, PMT = 1,000,000, I/Y = 5, FV = 0

CPT PV_{10}

↓

$PV_0 = {PV_{10}}/{(1.05)^{10}}$

N = 30, PMT = 1,000,000, I/Y = 5, FV = 0

CPT PV_9

↓

$PV_0 = {PV_9}/{(1.05)^9}$

→ Present value of a perpetuity: $PV = A/r$

- level CFs
- sequential
- infinite

Page 12
LOS c
- calculate
- interpret

e.g./ \$10 at t = 1 forever

r = 20%

$PV = 10/.20 = \$50$

e.g./ £ 100/yr. - perpetual, r = 5%

$PV = 100/.05 = £2,000$

e.g./

£100 100 100 100

t = 0 5 6 7 8

(r = 5%)

$PV_5 = 100/.05 + 100 = 2100$

$PV_0 = \frac{2100}{(1.05)^5} = 1645.40$

$PV_4 = 100/.05 = 2000$

$PV_0 = \frac{2000}{(1.05)^4} = 1645.40$

e.g./

long: 100 100 100 100 100 — t = 0 1 2 3 4 5 → perpetuity at t = 0

$PV_0 = 100/.05 = 2,000$

short: 100 — t = 0 5 → perpetuity at t = 4

$P_0 = {PV_4}/{(1.05)^4} = \frac{1645.40}{354.60}$

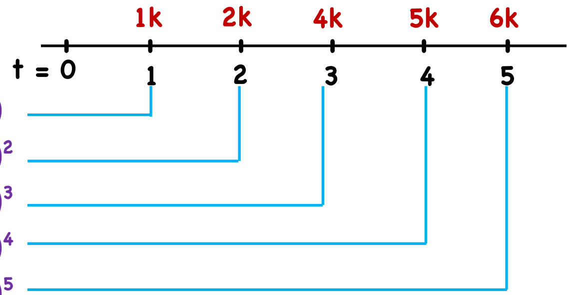
= 4 yr. ord. annuity

$$100\left[\frac{1-\frac{1}{(1.05)^4}}{.05}\right] = 354.60$$

N = 4
PMT = 100
I/Y = 5 CPT PV
FV = 0

→ **Present value of a series of unequal CFs:**

Page 13
LOS c
- calculate
- interpret



$1000/(1.05)$
$2000/(1.05)^2$
$4000/(1.05)^3$
$5000/(1.05)^4$
$6000/(1.05)^5$
$= 15{,}036.46$

Calculator:

2nd	CF	2nd	CE/C
CF_0		↓	
CO1	1000	ENTER	↓ ↓
CO2	2000	ENTER	↓ ↓
CO3	4000	ENTER	↓ ↓
CO4	5000	ENTER	↓ ↓
CO5	6000	ENTER	↓ NPV
I	5	ENTER	
NPV		CPT	

15,036.46

→ **Solve for r, N or PMT/**

Page 14
LOS c
- calculate
- interpret

$FV = PV(1 + r)^N$ → solve for r

$$FV/PV = (1+r)^N$$
$$(FV/PV)^{1/N} = 1 + r$$
$$r = (FV/PV)^{1/N} - 1$$

e.g./

Year	Sales	Profit
2008	10,503	822.5
2012	14,146.4	796.4

- growth in Sales ⇒ $14{,}146.40 = 10{,}503\,(1 + g)^4$

$$g = (14146.40/10503)^{1/4} - 1 = .07729 \quad (7.73\%)$$

- growth in Profit ⇒ $796.4 = 822.5\,(1 + g)^4$

$$g = (796.4/822.5)^{1/4} - 1 = -.00803 \quad (-.803\%)$$

e.g./ 2012 - 7.35M units sold
2007 - 8.52M units sold
find g

$$g = (FV/PV)^{1/N} - 1$$
$$= (7.35/8.52)^{1/5} - 1 = -.02911 \quad (-2.911\%)$$

Note: g is called the compound annual growth rate

Page 15
LOS c
- calculate
- interpret

→ Solve for r, N or PMT/

e.g./ How long will it take to double €10M at 7% compounded annually?

→ Solve for N: $FV = PV(1+r)^N$

$$(1+r)^N = {}^{FV}/_{PV}$$

$$N\ In(1+r) = In({}^{FV}/_{PV})$$

$$N = \frac{In({}^{FV}/_{PV})}{In(1+r)}$$

$$20M = 10M\,(1.07)^N$$

$$N = \frac{In({}^{20}/_{10})}{In(1.07)} = \frac{.69314}{.06765} = 10.244$$

calculator: 20 ÷ 10 = LN
÷ (1.07 LN) =

→ solve for PMT: $100,000 mortgage, 30 yrs., 8% compounded monthly

$$PV = A\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right]$$

$$A = {}^{PV}\Big/\left[\frac{1-\frac{1}{(1+r)^N}}{r}\right] = 100,000\Big/\left[\frac{1-\frac{1}{(1+.08/12)^{12\times30}}}{.08/12}\right] = \frac{100,000}{136.2834} = 733.76$$

N = 360
PV = -100,000
FV = 0
I/Y = 8/12 = .66$\dot{6}$

CPT PV

$\downarrow$ 666.66 interest

Page 16
LOS d
- demonstrate

retirement income

t = 0	1	2	15	16	40	41	42	60
age 22	2k	2k	2k			100k (age 63)	100k	100k

$\underbrace{2k \quad 2k \quad 2k}_{\text{savings}}$

how much to save each yr.? Assume r = 8%
(PMT = ?)

1. FV_{15} → N = 15, PMT = 2000, I/Y = 8, PV = 0 CPT FV = 54,304.23
2. PV_{40} → N = 20, PMT = 100,000, I/Y = 8, FV = 0 CPT PV = 981,814.74
3. PV_{15} → 143,362.53 - 54,304.23 = 89,058.30
N = 25, FV = 0, I/Y = 8, PV = -89,058.30
CPT PMT = 8,342.87

$$PV_{15} = {}^{PV_{40}}\Big/_{(1.08)^{25}} = 143,362.53$$

or/

$PV_{40} = PV_{15}(1.08)^{25} = 371,901.17$
$FV_{40} = 981,814.74 - 371,901.17 = 609,913.56$

N = 25, FV = 609,913.56, I/Y = 8
PV = 0 CPT PMT = 8342.87

Organizing, Visualizing, and Describing Data

a. identify and compare data types

b. describe how data are organized for quantitative analysis

c. interpret frequency and related distributions

d. interpret a contingency table

e. describe ways that data may be visualized and evaluate uses of specific visualizations

f. describe how to select among visualization types

g. calculate and interpret measures of central tendency

h. evaluate alternative definitions of mean to address an investment problem

i. calculate quantiles and interpret related visualizations

j. calculate and interpret measures of dispersion

k. calculate and interpret target downside deviation

Organizing, Visualizing, and Describing Data

Page 1
LOS a
- identify
- compare

Data → a collection of numbers, characters, words, or text that represent facts or information

I/ Numerical or Categorical
(quantitative) (qualitative)

Categorical: values that describe a quality or characteristic
- mutually exclusive labels or groups

(N) c) nominal → no logical order (e.g. sectors of the economy)
(O) d) ordinal → has a logical order or rank
- no information in the distance between groups however

Numerical: measured or counted quantities
integer (I) a) Discrete → limited to a finite number of values
ratio (R) b) Continuous → can take on any value within a range

Ex. #1

Page 2
LOS a
- identify
- compare

II/ Cross-sectional vs. Time-series vs. Panel Data

Definitions: variable → a particular quality or characteristic
(e.g. stock price, height)
observation → value of a specific variable
(e.g. GM at \$53.30, Tom at 96 kg)

a) cross-sectional - multiple observations of a particular variable
(stock prices of 60 companies)
b) time-series - multiple observations of a particular variable for the same observational unit over time
(GM's stock price over the last 60 months)
c) panel data - cross-sectional + time-series

	GM	F . . . TSLA	CS	
	$month_t$			
	$month_{t-1}$	data		panel
TS	$month_{t-2}$	table		data
	⋮			

Page 3
LOS a
- identify
- compare

III/ Structured vs. Unstructured Data

a) Structured → highly organized in a pre-defined manner (e.g. stock prices, returns, EPS)

b) Unstructured → no organized form (news, social media posts, company filings, audio/video)

- also called 'alternative data'
  - produced by individuals
  - generated by business processes (credit card transactions)
  - generated by sensors
- to be useful in data analysis → must be transformed into structured data

Ex. #2

Page 4
LOS b
- describe

One-dimensional array (1 variable)
- e.g. a column of a spreadsheet (CS or TS)

Two-dimensional rectangular array (two or more variables)
- data table (CS or panel)

	Var. 1	Var. 2	...	Var. N
Comp. 1	-	-		-
Comp. 2	-			
⋮				
Comp. m	-			

LOS c
- interpret

- Frequency distribution (one-way table)
  - the number of observations of a specific value or group of a variable
  - sorted in ascending or descending order

Exhibit #8

Page 5
LOS c
- interpret

- **Absolute frequency** - actual count of observations per value of the variable ($\sum = N$)
- **Relative frequency** - %'age of observations per value of the variable (abs. freq./total N) $\sum = 100\%$

- for numerical data: create non-overlapping intervals (bins)
  - sort data in ascending order
  - find the range: max. - min.
  - decide on the number of intervals (k)
  - interval width = range/k (round up always)

too few = too much aggregation - loss of info.
too many - not enough aggregation - too much noise

Interval 1 = min. value + width

e.g. [0,5) , [5,10) , [10,15] (k = 3)

$0 \leq x < 5$ $10 \leq x \leq 15$

- each obs. falls into only one interval

Page 6
LOS c
- interpret

e.g./

-4.57
-4.04
-1.64
0.28
1.34
2.35
2.38
4.28
4.42
4.68
7.16
11.43

N = 12

1 ascending order

2 Range = 11.43 - (-4.57) = 16

3 let k = 4

4 width = range/k = 16/4 = 4

5 Intervals:

[-4.57, -0.57)	[-0.57, 3.43)	[3.43, 7.43)	[7.43 - 11.43]
3	4	4	1

(6 → counts)

- cumulative — absolute frequency / relative frequency } a sequence of partial sums that sum to N or 100%

Exhibit #11

Page 7
LOS d
- interpret

• Contingency table - summarizes data for 2 or more categorical variables
(helps visually find patterns)
• 2-way table = 2 variables

Exhibit 14 Portfolio Frequencies by Sector and Market Capitalization

Sector Variable (5 Levels)	Market Capitalization Variable (3 Levels) Small	Mid	Large	Total
Communication Services	55	35	20	110
Consumer Staples	50	30	30	110
Energy	175	95	20	290
Health Care	275	105	55	435
Utilities	20	25	10	55
Total	575	290	135	1,000

Var 2
Var. 1
Rows = 5
Columns = 3
R x C table
(5 x 3)
row totals
- marginal frequencies
N
joint frequency
column totals
- marginal frequencies
Exhibit 15/16 → (can be abs. or %'age)

Page 8
LOS d
- interpret

• applications/

1/ confusion matrix

Exhibit 17 Confusion Matrix for Bond Default Prediction Model

Predicted Default	Actual Default Yes	No	Total
Yes	300	40	340
No	10	1,650	1,660
Total	310	1,690	2,000

2/ potential association between 2 categorical variables
- use of a 'chi-square test of independence'

Actual
Observed Values

	Low Risk	High Risk	
Growth	73	26	99
Value	183	33	216
	256	59	315

Expected
Observed Values

	Low Risk	High Risk	
Growth	80.457	18.543	99
Value	175.543	40.457	216
	256	59	315

$$\frac{(99 \times 256)}{315}$$

$$\chi^2 = \sum \frac{(0-e)^2}{e}$$

df = (C - 1)(R - 1)

$$= \frac{(73-80.457)^2}{80.457} + \frac{(183-175.543)^2}{175.543} + \frac{(26-18.543)^2}{18.543} + \frac{(33-40.457)^2}{40.457}$$

$= 5.38$

df = (2 - 1)(2 - 1) = 1

Page 9
LOS e
- describe
- evaluate

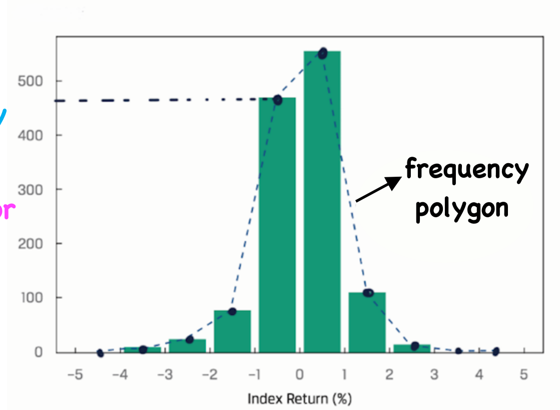
Visualization: presentation of data in pictorial or graphical format

1/ Histogram & Frequency Polygon

represents the distribution of numerical data

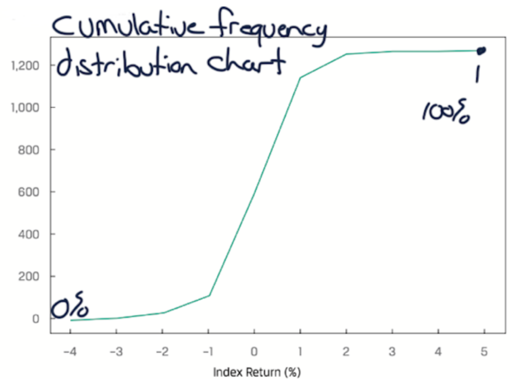
y-axis: frequency

(can be absolute or relative)



x-axis → intervals/values

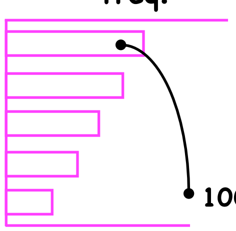
Cumulative frequency distribution chart



Page 10
LOS e
- describe
- evaluate

2/ Bar Chart - represent the frequency distribution of categorical data

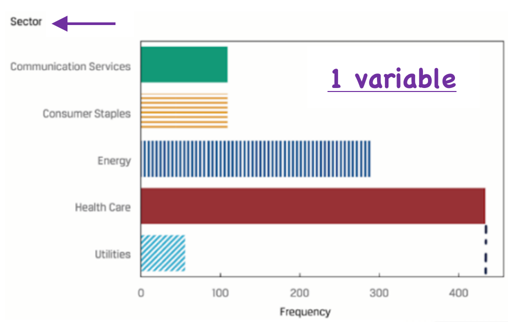
freq.



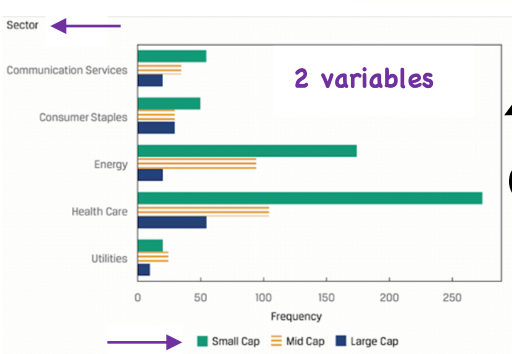
100%

%

Pareto chart



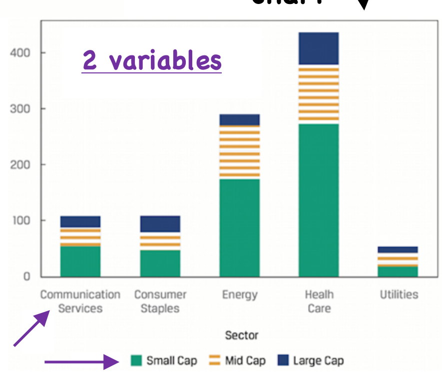
horizontal
- can be vertical



grouped bar chart

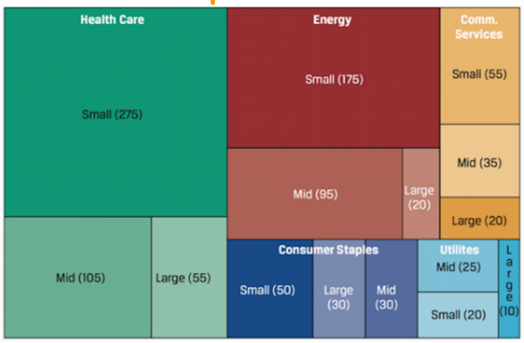
(a.k.a. clustered bar chart)

Stacked bar chart



Page 11
LOS e
- describe
- evaluate

3/ Tree-Map



- a set of coloured rectangles to represent groups
- area = %'age of group
- green = health care (1 category)
  → nested rectangles → market cap (other category)
→ within each market cap:
  → more nested rectangles for another category

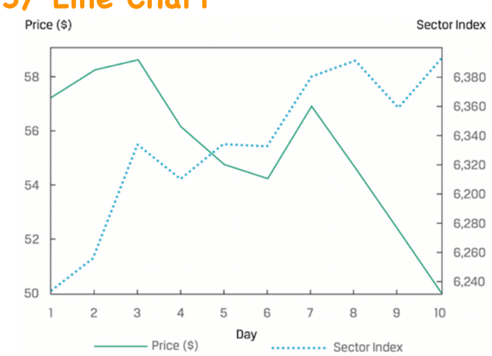
4/ Word Clouds (a.k.a. tag cloud)

- depicts frequency of unstructured data (e.g. text)

→ size of each word proportional to its frequency in the text
- colour can be used to display different sentiment

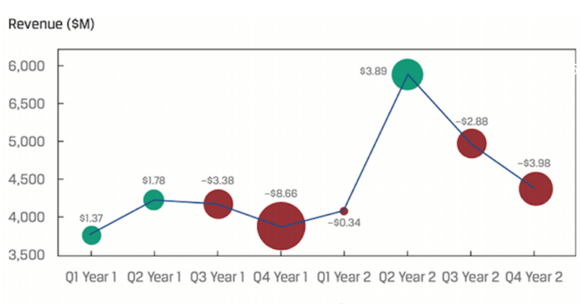
Page 12
LOS e
- describe
- evaluate

5/ Line Chart



- used to visualize ordered observations
- typically used for time series data
- facilitates showing changes and underlying trends (aids in forecasting)

← can show more than one time series

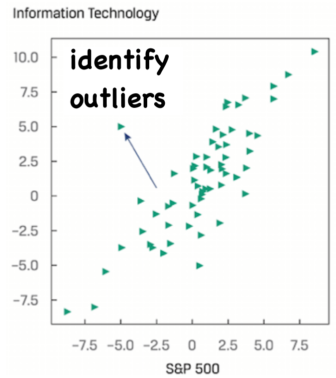


- adding a third characteristic
  (revenue + time = line chart)
  (rev. + time + EPS = bubble line chart)

EPS^+ = green EPS^- = red

Page 13
LOS e
- describe
- evaluate

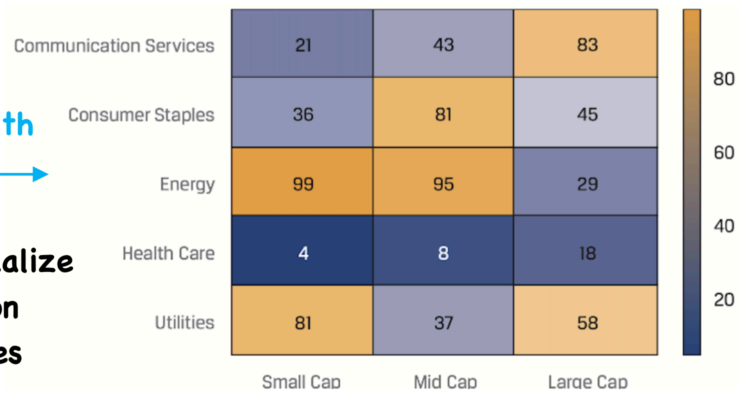
6/ Scatter Plot



- used to visualize the joint variation in 2 numerical values
  - may be no relationship, a linear or non-linear relationship
- scatter plot matrix
  - assess for pairwise association among many variables **(Exhibit 32)**

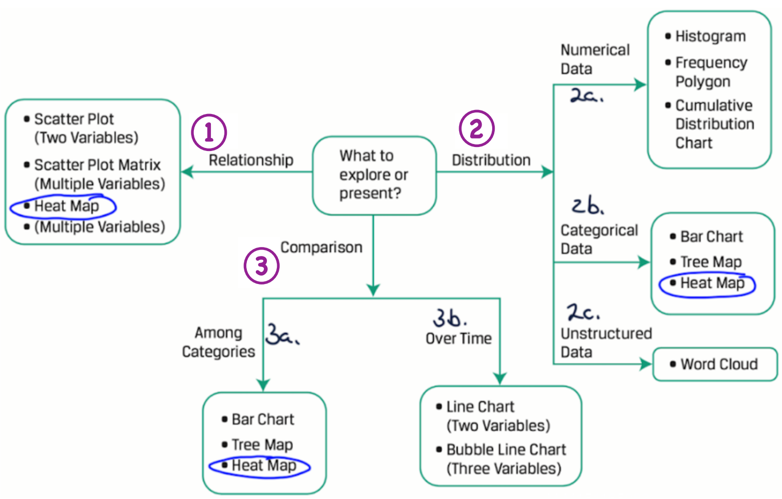
7/ Heat Maps

- contingency table with colour-coded cells →



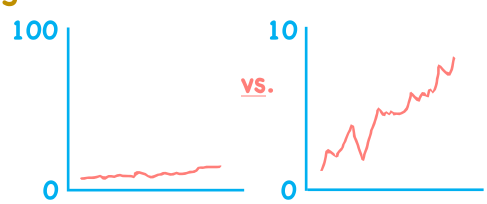
- can also be used to visualize the degree of correlation among different variables

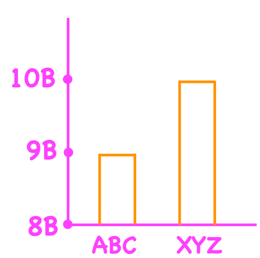
Page 14
LOS f
- describe



Pitfalls/

1. selecting an improper chart type - hinders accurate interpretation of data
2. Selecting data that favours a conclusion
3. truncating the range
4. extending the range

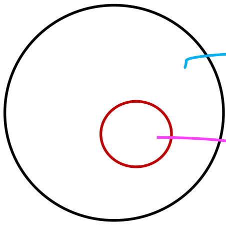




implies XYZ is 2x that of ABC

Page 15
LOS g
- calculate
- interpret

- measures of central tendency - specifies where data are centered
  (arithmetic mean, median, mode, weighted mean, geometric mean, harmonic mean)
- measures of location - deciles, quantiles, quintiles



population → parameters μ , σ

measures of dispersion

sample → sample statistics $\underbrace{\bar{X} \text{ , } S}_{\text{descriptive statistics}}$

1/ Arithmetic mean

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

↓
X-bar

- average sales of 50 companies } cross-sectional mean
- average sales for last 10 yrs. for GM } time-series mean

Page 16
LOS g
- calculate
- interpret

1/ Arithmetic mean

property: $\sum_{i=1}^{n}(X_i - \bar{X}) = 0$

- deviations from the mean indicate risk
  (variance, skew, kurtosis)

disadvantage: sensitive to outliers

e.g. 1, 2, 3, 4, 5, 6, 1000 mean = $^{1021}/_{7}$ = 145.86 → not representative of any value

Options

1/ Do nothing - appropriate if the value is legitimate and correct
- may contain meaningful information

2/ Delete - trimmed mean → exclude a small %'age of lowest and highest values

e.g. 5% trimmed mean:
- deletes top 2.5% and bottom 2.5%

Page 17
LOS g
- calculate
- interpret

1/ Arithmetic mean
Options
3/ replace with another value
95% winsorized mean → top 2.5% of values replaced by the value at which all others lie above (opposite for the bottom)

e.g./ 100
88 — 4 obs. (2.5%) - all 4 assigned 88
(25 - 75) ← majority
12 — 4 obs. (2.5%) - all 3 assigned 12
0

2/ Median - middlemost value of a set of observations

odd # of obs. median = $(n+1)/2$ e.g. n = 11 median = $(11+1)/2 = 6^{th}$ obs.

even # of obs. median = $\frac{\frac{n}{2}+\frac{(n+2)}{2}}{2}$ e.g. n = 10 median = $\frac{\frac{10}{2}+\frac{12}{2}}{2} = \frac{\text{5th + 6th obs.}}{2}$

- not affected by extreme values (outliers)
  - useful for describing central tendency for a non-symmetrical distribution

Page 18
LOS g
- calculate
- interpret

3/ Mode - the most frequently occurring value in a distribution
- unimodal → only 1 value that is most frequent
- bi-modal → two values have the highest frequency
- tri-modal → three …

or no mode → no value occurs more frequently than any other value (uniform distribution)

- only measure of central tendency that can be used with nominal data
- For a symmetrical distribution: mode = median = mean

4/ Weighted mean

$\bar{X}_w = \sum_{i=1}^{n} w_i x_i$ where $\sum_{i=1}^{n} w_i = 1$

X-bar sub-w

- common in finding R_P or $E(R_P)$

$R_P = W_A R_A + W_B R_B + \ldots + W_N R_N$

$E(R_P) = W_A E(R_A) + W_B E(R_B) + \ldots + W_N E(R_N)$

$W_i > 0$ = long position
$W_i < 0$ = short position

Page 19
LOS g
- calculate
- interpret

4/ Weighted mean - weights can be probabilities

$R_{SP500} = P_A \cdot R_A + P_B \times R_B + P_C R_C$ where $\sum P = 1$

P_A: bullish, P_B: neutral, P_C: bearish

5/ Geometric mean → used with rates of change over time or to compute growth rates

$\bar{X}_G = \sqrt[n]{X_1 X_2 X_3 \ldots X_n}$ $\quad X_i \geq 0$ for all $i = 1, 2, 3 \ldots n$ → e.g./ $(8 \times 9 \times 10)^{\frac{1}{3}} = 8.962$

or $\ln(\bar{X}_G) = \dfrac{\ln(X_1 X_2 X_3 \ldots X_n)}{n} \longrightarrow \bar{X}_G = e^{\ln(\bar{X}_G)} \longrightarrow e^{\ln(720)/3} = 8.962$

$$\bar{X}_G = R_G = \left[\prod_{t=1}^{T}(1 + R_T)\right]^{\frac{1}{T}} - 1$$

(* critical to know)

- also referred to as compounded returns

Property:

$\bar{X}_G \leq \bar{X}_A$

- difference between them grows as variability increases
(example #10)

Page 20
LOS g
- calculate
- interpret

5/ Geometric mean

e.g. #1/

YR1	YR2	YR3
7.8%	6.3%	-1.5%

$$\bar{X}_A = \frac{7.8\% + 6.3\% - 1.5\%}{3}$$

$$= 12.6\%/3 = 4.2\%$$

- but $(1.042)^3 - 1 = 13.137\%$

$1 invested would actually grow to: $(1.078)(1.063)(.985) = 1.128725$ (12.8725%)

$$\bar{X}_G = (1.128725)^{\frac{1}{3}} - 1 = 4.119\%$$

$(1.04119)^3 = 1.128725$

e.g. #2/ Beg. Sales $12M
Ending Sales $21M
N = 6 yrs.

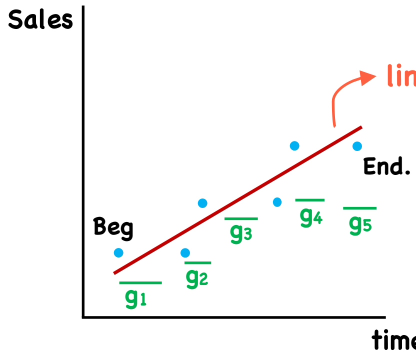
$$\left[1 + \left(\frac{21 - 12}{12}\right)\right]^{\frac{1}{6}} - 1 = 9.775\%$$

$\$12M(1.09775)^6 = \$21M$

Page 21
LOS g
- calculate
- interpret

5/ Geometric mean

e.g. #3/



$$[(1+g_1)(1+g_2)(1+g_3)(1+g_4)(1+g_5)]^{\frac{1}{5}}-1$$

- $\bar{X}_G$ → growth rate of an investment (constant/yr.) over multiple periods
- $\bar{X}_A$ → average single period return

∴ forecast of returns (or growth rate) in one YR? → $\bar{X}_A$
over multiple periods? → $\bar{X}_G$

$$\bar{X}_G = \bar{X}_A - \sigma^2/2$$

Page 22
LOS g
- calculate
- interpret

6/ Harmonic Mean

$$\bar{X}_H = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{X_i}\right)}$$

- gives much less weight to outliers

- appropriate for averaging ratios when the ratios are repeatedly applied to a fixed quantity to yield a variable number of units

e.g. dollar cost averaging

invest €1,000 a month for 2 months

m_1 → €10/sh. m_2 → €15/sh.

$1000/_{10} = 100$ sh.

$1000/_{15} = \frac{66.67\text{ sh.}}{166.67}$

∴ $2000/_{166.67} = 12/\text{sh.}$

or/ $\bar{X}_H = \frac{2}{\frac{1}{10}+\frac{1}{15}} = \frac{2}{.1+.067} = 12$

Page 23
LOS h
- select

$\overline{X}_A \times \overline{X}_H = \overline{X}_G^2$ and $\overline{X}_A > \overline{X}_G > \overline{X}_H$

$\overline{X}_A$: when including all values, including outliers

$\overline{X}_G$: when compounding is involved

$\overline{X}_H$: to avoid outliers

LOS i
- calculate
- interpret

Quantiles/

Quart. 25%, 50%, 75% → interquartile range: 1QR = $Q_3 - Q_2$

Quint. 20%, 40%, 60%, 80%

Dec. 10%, 20%, 30% … 90%

Percent 1%, 2% . . . 99% → $L_y = (n+1)\frac{y}{100}$ (with data in ascending order)

location

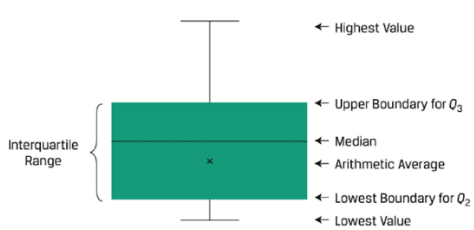
• an n ↑, Ly becomes more accurate

Ly is an integer → done

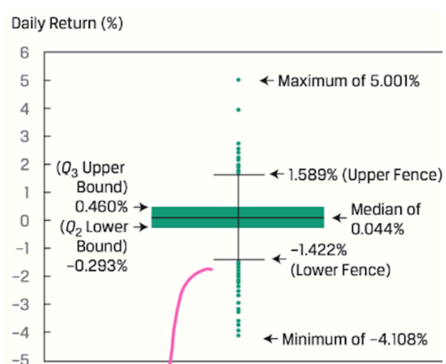
Ly is a decimal → interpolation e.g. 6.8 X_6 — X_7 → .8

$X_6 + (X_7 - X_6).8$

Page 24
LOS i
- calculate
- interpret



Box & whisker plot



upper fence = (1.5 x 1QR) + upper bound

lower fence = lower bound − (1.5 x 1QR)

uses/ • rank performance of portfolios and investment managers in terms of percentile/quartile in which they fall

• investment research → bottom return decile → short
→ top return decile → long

short, long } long/short HF

(more at L3)

Page 25
LOS j
- calculate
- interpret

Dispersion → the variability around the central tendency

- measures of absolute dispersion

1/ Range = max. value - min. value (56 - 12 = 44)
or max. value to min. value (ranges from 56 to 12)

- uses only 2 observations
- tells us nothing about the shape of the distribution however

2/ Mean Absolute Deviation (MAD)

$$\text{MAD} = \frac{\sum_{i=1}^{n}|X_i - \bar{X}|}{n}$$ - uses all the observations

3/ Variance and standard deviation
(S^2 or σ^2) (S or σ)

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1} \longrightarrow \text{df}$$

Let n = 10. If we know $\bar{X}$, we can only take 9 X_i at random, the 10th is constrained.
∴ we do not have n independent variables, we have $n-1$

Page 26
LOS j
- calculate
- interpret

3/ Variance and standard deviation

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$ - variance of $\bar{X}_A$

measured in units squared

e.g. $16(\%^2)$
$\sqrt{16\%^2} = 4\%$

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$$ - sd of $\bar{X}_A$

expressed in the same units of measurement as the mean

- for $\bar{X}_G$, recall that $\bar{X}_G \approx \bar{X}_A - {S^2}/{2}$

$S_G^2 = S^2 t$

$S_G = S\sqrt{t}$

application: BSM

$$Z = \frac{X_i - \bar{X}}{S}$$

$$Z = \frac{\ln\left({S_T}/{X}\right) - \left(r - {\sigma^2}/{2}\right)t}{\sigma\sqrt{t}}$$

(Level 2 lookahead)

Page 27
LOS k
- calculate
- interpret

Target Downside Deviation - only concerned with downside risk

- Target Semideviation → a measure of dispersion below the target

$$S_{target} = \sqrt{\frac{\sum^{n}_{\forall X_i < B}(X_i - B)^2}{n - 1}}$$

↓

full n, not just n of $X_i < B$

e.g. 10
8
6 → Let B = 5
4
2
0

(5 - 5)² + (5 - 5)²
+ (5 - 5)² + (4 - 5)²
+ (2 - 5)² + (0 - 5)²

- as $B \uparrow$, $S_{target} \uparrow$

(example 18 and 19)

Coefficient of Variation → measure of relative dispersion

$CV = {}^{S}/_{\bar{X}}$ where $\bar{X} > 0$

e.g./ for returns, CV measures the risk per unit of return

Page 28
LOS k
- calculate
- interpret

Coefficient of Variation

$CV = {}^{S}/_{\bar{X}}$ - allows for direct comparisons of dispersion across different data sets

e.g./ $\bar{X} = 20$, $S = 6$ vs. $\bar{X} = 1000$, $S = 190$ } which one has greater dispersion?

$CV = {}^{6}/_{20} = .3$ $CV = {}^{190}/_{1000} = .19$ } lower = less dispersion

Probability Concepts

a. define a random variable, an outcome, and an event

b. identify the two defining properties of probability, including mutually exclusive and exhaustive events, and compare and contrast empirical, subjective, and a priori probabilities

c. describe the probability of an event in terms of odds for and against the event

d. calculate and interpret conditional probabilities

e. demonstrate the application of the multiplication and addition rules for probability

f. compare and contrast dependent and independent events

g. calculate and interpret an unconditional probability using the total probability rule

h. identify the most appropriate method to solve a particular counting problem and analyze counting problems using factorial, combination, and permutation concepts

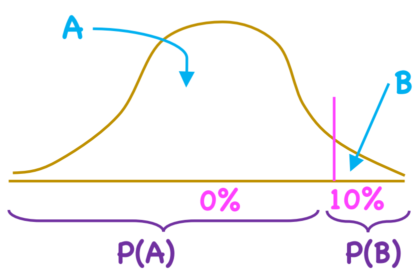
Probability Concepts

Page 1
LOS a
- define

- Random variable - a quantity whose future outcomes are uncertain (e.g. returns)

- Outcome - a possible value of a random variable (e.g. 4.3%)

- Event - a specified set of outcomes e.g.: A = (r_P < 10%) B = ($r_P \geq$ 10%)



$$P(A) + P(B) = 100\%$$

$$\left(\sum_{i=1}^{n} P_i = 1\right)$$

- if an event is impossible: P(E) = 0%
- if an event is certain: P(E) = 100%

} not really random

Page 2
LOS b
- identify
- compare
- contrast

- Property 1: $0 \leq P(E) \leq 1$
- Property 2: $\sum_{i=1}^{n} P_i(E_i) = 1$

} where E are
1/ mutually exclusive
(if one happens, another can't)
2/ exhaustive
(covers all possible outcomes)

- How are probabilities estimated?

1/ empirical probabilities → based on historical observation
- past is assumed to be representative of the future
- historical period must include occurences of the event

$$P(A) = \frac{\text{absolute frequency of event E}}{\text{total frequency of all events}} = \text{relative frequency of event A}$$

2/ subjective probabilities → adjust an empirical probability based on intuition or experience
- when there is a lack of empirical observations
- to make a personal assessment

Page 3
LOS b
- identify
- compare
- contrast

- How are probabilities estimated?
  3/ a priori probabilities - arriving at a conclusion based on deductive reasoning
  e.g. P(1) = 1/6 (roll a die, get a 1)

LOS c
- describe

expressed as

- Odds for: $\frac{P(E)}{1-P(E)}$ a to b probability = $\frac{a}{a+b}$
- Odds against: $\frac{1-P(E)}{P(E)}$ b to a probability = $\frac{b}{a+b}$

e.g./
P(E) = 1/8 → 1 to 7 - for each occurrence of E, we expect 7 non-occurrences

P(A) = 3/17 → $\underbrace{\text{3 to 14}}_{\text{odds for}}$ - for every 3 occurrences of A, we expect 14 non-occurrences

Page 4
LOS c
- describe

- from odds to probability: for: 1 to 4 → $\frac{1}{1+4} = .2$

  against: 4 to 1 → $\frac{4}{1+4} = .8$

e.g./ Wager:
A = win, B = loss } mutually exclusive, exhaustive

Odds for = $\frac{1/16}{15/16}$ → 1 to 15

$1 bet pays $16, profit = $15

P(A) = 1/16, P(B) = 15/16 } $0 \leq P(E) \leq 1$, $\sum P_i(E) = 1$

Odds against $\frac{15/16}{1/16}$ → 15 to 1

lose $1

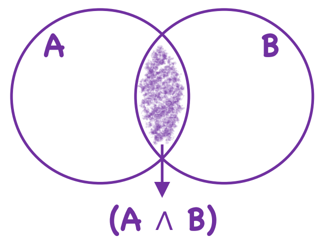
∴ expected profit = $\left(1/16\right)15 + \left(15/16\right)(-1) = 15/16 - 15/16 = 0$

Page 5
LOS d
- calculate
- interpret

$P(A)$ → unconditional probability
$P(A|B)$ → conditional probability (prob. A given B)

$P(A|B) = \frac{P(AB)}{P(B)}$ → called a joint probability
• probability of A and B occurring

e.g./ $P(B) = .5$
$P(AB) = .1$
$\therefore P(A|B) = {.1}/{.5} = 20\%$



LOS e
- demonstrate

Multiplication Rule: $P(AB) = P(A|B)P(B)$

A = YR1 Winner
A^c = YR1 Loser
B = YR2 Winner
B^c = YR2 Loser

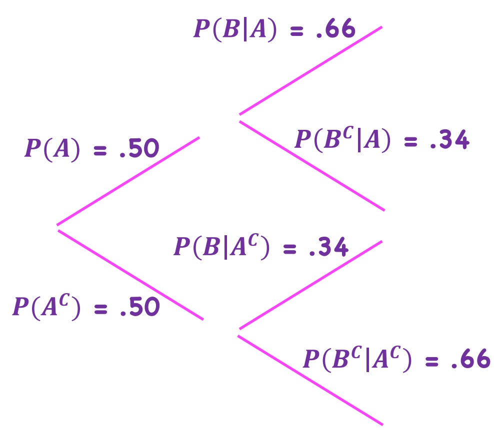
Exhibit 3 Persistence of Returns: Conditional Probability for Year 2 Performance Given Year 1 Performance

	Year 2 Winner	Year 2 Loser
Year 1 Winner	66% $P(B|A)$	34% $P(B^c|A)$
Year 1 Loser	34% $P(B|A^c)$	66% $P(B^c|A^c)$

ex. #3

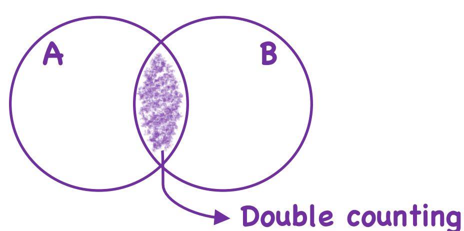
Page 6
LOS e
- demonstrate

Tree/
1/ YR1 Winner = A
YR1 Loser = A^c
YR2 Winner = B
YR2 Loser = B^c



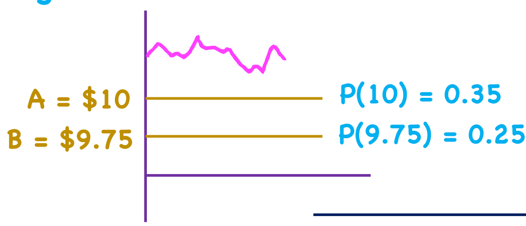
calculate: $P(A^CB^C)$
$= P(B^C|A^C)P(A^C)$
$= .66 \times .5$
$= 0.33$

Additional Rule: P(A or B) = $P(A \vee B) = P(A) + P(B) - P(AB)$



Page 7
LOS e
- demonstrate

e.g./



find: P(A or B)
P(A) + P(B) – P(AB)
= .35 + .25 – .25
= .35

LOS f
- compare
- contrast

- **Independent event:** 2 events are independent iff

$P(A|B) = P(A)$ → knowing B tells us nothing about A
or $P(B|A) = P(B)$

∴ $P(AB) = P(A)P(B)$ e.g. $P(H|H) = P(H) = 50\%$

- **Dependent event:** P(A) is related to P(B)

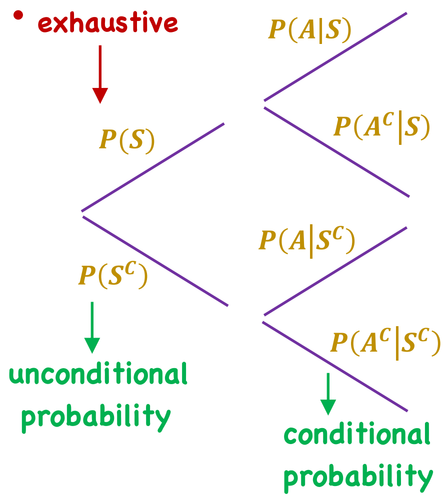
e.g. A = stock ABC rises
B = SnP500 rises
} A is most likely dependent upon B

ex. #6/7

Page 8
LOS g
- calculate
- interpret

Total Probability Rule: $P(A) = P(A|S)P(S) + P(A|S^C)P(S^C)$

- when the conditioning events are
- mut. excl.
- exhaustive



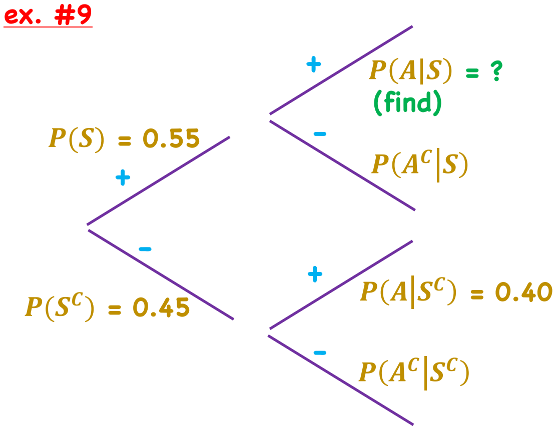
$P(A) =$
$P(A|S)P(S)$
$+$
$P(A|S^C)P(S^C)$

$P(A^C) =$
$P(A^C|S)P(S)$ → multiplication rule
$+$
$\underbrace{P(A^C|S^C)P(S^C)}_{\text{total probability rule}}$

Page 9

LOS g
- calculate
- interpret

ex. #9



$P(A)$ = 0.55 $\quad$ $P(A^C)$ = 0.45

= $P(A|S)$.55

+

$\underline{.40(.45)}$ → .55 = .55X + .18

.37 = .55X

X = $\frac{.37}{.55}$ = .6727

LOS h
- identify
- analyze

Counting/
1/ Multiplication

e.g./ Portfolio subdivided by
- Domestic/Foreign
- then by 4 industries
- then by 3 size categories

- how many sub-portfolios?

2 x 4 x 3 = 24

Counting/

LOS h
- identify
- analyze

2/ Factorial

$n!$

e.g./ 3 analysts to cover 3 industries

3 x 2 x 1 = 3!

3/ Multinomial – the number of ways that n objects can be labelled with k labels with n_1 to n_k representing the size of each label category

$$\frac{n!}{n_1!\,n_2!\ldots n_k!}$$

e.g./ Rank 18 funds by total return
- each 18 assigned to 5 categories

high risk	above-avg. risk	avg. risk	below-avg. risk	low risk
4	4	3	4	3

ABCD is the same as ACBD

Factorial of 4!

$$\therefore \frac{18!}{4!\,4!\,3!\,4!\,3!}$$

Counting/

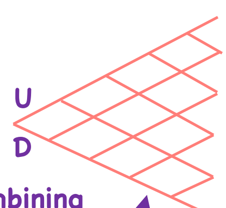
LOS h
- identify
- analyze

if k = 1 → factorial
if k = 2 → combination or permutation

4/ Combination

$${}_{n}C_{r} = \binom{n}{r} = \frac{n!}{(n-r)!\,r!}$$

- the number of ways of selecting r objects from n where order does not matter



recombining lattice

e.g./ For 5 price changes with 3 Us, how many ways can this occur?

$${}_{5}C_{3} = \frac{5!}{(5-3)!\,3!} = \frac{5 \times 4 \times 3!}{2!\,3!} = 10$$

5/ Permutation – if k = 2 and order matters

$${}_{n}P_{r} = \frac{n!}{(n-r)!}$$

Ex 17, 17, 19

Common Probability Distributions

a. define a probability distribution and compare and contrast discrete and continuous random variables and their probability functions

b. calculate and interpret probabilities for a random variable given its cumulative distribution function

c. describe the properties of a discrete uniform random variable, and calculate and interpret probabilities given the discrete uniform distribution function

d. describe the properties of the continuous uniform distribution, and calculate and interpret probabilities given a continuous uniform distribution

e. describe the properties of a Bernoulli random variable and a binomial random variable, and calculate and interpret probabilities given the binomial distribution function

f. explain the key properties of the normal distribution

g. contrast a multivariate distribution and a univariate distribution, and explain the role of correlation in the multivariate normal distribution

h. calculate the probability that a normally distributed random variable lies inside a given interval

i. explain how to standardize a random variable

j. calculate and interpret probabilities using the standard normal distribution

k. describe the properties of the Student's *t*-distribution, and calculate and interpret its degrees of freedom

l. describe the properties of the chi-square distribution and the *F*-distribution, and calculate and interpret their degrees of freedom

Common Probability Distributions

Page 1
LOS a
- define
- compare
- contrast

- **Probability distribution** → specifies the probabilities associated with the possible outcomes of a random variable
  (uniform, binomial, normal, lognormal, Student's t, chi-square, F-distribution)
- **Random variable** → a quantity whose future outcomes are uncertain
  - discrete – take on at most a countable number of possible values (possibly infinite)
  - continuous – cannot count the possible values
- every random variable is associated with a probability distribution that describes the variable completely
- **Probability function** → specifies the probabilities that a random variable can take i.e. P(X = x)
  - discrete variables: p(x)
  - continuous variables: f(x) → the probability density function (pdf)

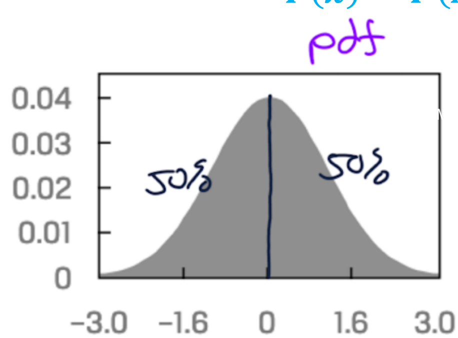
Page 2
LOS a
- define
- compare
- contrast

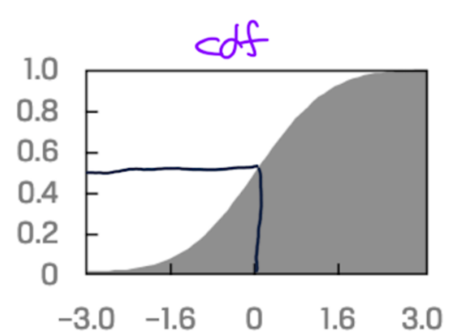
- **Probability function**
  - has 2 key properties
    1/ $0 \leq p(x) \leq 1$
    2/ $\sum p(x)$ over all values of X equals 1

LOS b
- calculate
- interpret

- **Cumulative distribution function (cdf)** → gives the probability that a variable X is less than or equal to a particular value x

$$F(x) = P(X \leq x)$$





Page 3
LOS c
- describe
- calculate
- interpret

• **Discrete uniform distribution**

A. Probability Function and Cumulative Distribution Function for a Discrete Uniform Random Variable

$X = x$	Probability Function $p(x) = P(X = x)$	Cumulative Distribution Function $F(x) = P(X \le x)$
1	0.125	0.125
2	0.125	0.250
3	0.125	0.375
4	0.125	0.500
5	0.125	0.625
6	0.125	0.750
7	0.125	0.875
8	0.125	1.000

$p(x) = \frac{1}{8}$ = .125

$P(X \le 7) = 0.875$

(P(1) + P(2) + … + P(7) = 7 x 0.125
= .875)

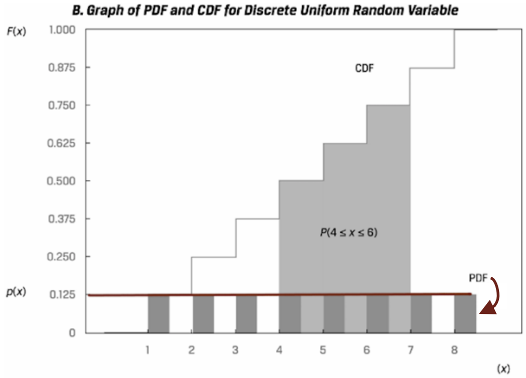
$P(4 \le X \le 6)$ = P(4) + P(5) + P(6)
= 3 x .125 = .375

or/ $P(x \le 6) - P(x \le 3)$
$= 0.750 - 0.375 = 0.375$

$P(4 < x \le 6)$ = P(5) + P(6) = 0.250

or/ $P(x \le 6) - P(x \le 4)$
$= 0.750 - 0.500 = 0.250$

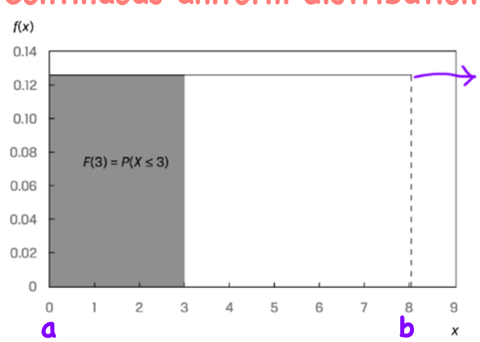
B. Graph of PDF and CDF for Discrete Uniform Random Variable



Page 4
LOS d
- describe
- calculate
- interpret

• **Continuous uniform distribution**

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$



$$f(x) = \frac{1}{8-0} = \frac{1}{8} = .125$$

$$F(3) = P(x \le 3) = \frac{x-a}{b-a} = \frac{3-0}{8-0} = \frac{3}{8} = .375$$

example #2/

LOS e
- describe
- calculate
- interpret

• **Bernoulli random variable:** the outcome of a trial that produces one of two outcomes (1 or 0)

$p(1) = p$ where p = success

$p(0) = 1 - p$

• in n trials, we can have 0 to n successes
- if each trial is a random var., then the # of successes in n trials is also a random var

Page 5
LOS e
- describe
- calculate
- interpret

• Binomial Random Variable – # of successes in n Bernoulli trials.

• assumptions: 1/ p is constant for all trials
2/ trials are independent

– a binomial random variable has a distribution completely described by 2 parameters $x \sim B(n, p)$

Q1: how many successes (x) are in n trials?

– does order matter? SSFF, SFSF, SFFS ? no. $\therefore\ {}_nC_r = \dfrac{n!}{(n-x)!\,x!}$

Q2: how probable is it to have x successes in n trials?

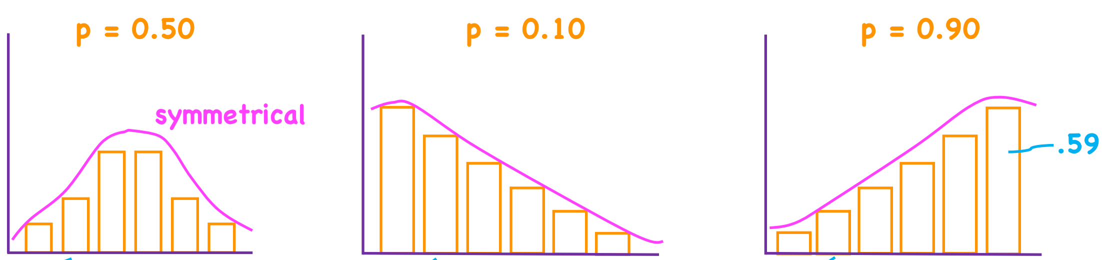
e.g.: P(SSFF) = $P(S)P(S)P(F)P(F)$

= $p \bullet p \bullet (1-p) \bullet (1-p) = p^2(1-p)^2 \rightarrow p^x(1-p)^{n-x}$

Page 6
LOS e
- describe
- calculate
- interpret

$$p(x) = \underbrace{\frac{n!}{(n-x)!\,x!}}_{\#\text{ of ways}} \bullet \underbrace{p^x(1-p)^{n-x}}_{\text{probability}}$$

n = 5



$\dfrac{5!}{(5-1)!\,1!} \bullet .5^5 = .15625$

$\dfrac{5!}{(5-1)!\,1!} \bullet 1^1(.9)^4 = .32805$

$\dfrac{5!}{(5-4)!\,1!} \bullet 9^1(.1)^4 = .00045$

example #3.

Page 7
LOS e
- describe
- calculate
- interpret

	Profitable	Losing	
BB001	3	9	12
BB002	5	3	8

$$\frac{n!}{(n-x)!\,x!}\, p^x(1-p)^{n-x}$$

3 A/ $P(X \leq 3)$

$$P(3) = \frac{12!}{(12-3)!3!} \bullet 5^3 \bullet 5^9 = .0537$$

$$P(2) = \frac{12!}{(12-2)!2!} \bullet 5^{12} = .01611$$

$$P(1) = \frac{12!}{(12-1)!1!} \bullet 5^{12} = .00293$$

$$P(0) = \frac{12!}{(12-0)!0!} \bullet 5^{12} = .000244$$

.072998
or 7.3%

B/ $P(X \geq 5)$

$$P(5) = \frac{8!}{(8-5)!5!} \bullet 5^8 = .21875$$

$$P(6) = \frac{8!}{(8-6)!6!} \bullet 5^8 = .109375$$

$$P(7) = \frac{8!}{(8-7)!7!} \bullet 5^8 = .03125$$

$$P(8) = \frac{8!}{(8-8)!8!} \bullet 5^8 = .003906$$

.363281
or 36.3%

Page 8
LOS e
- describe
- calculate
- interpret

	Mean	Variance
Bernoulli	p	p(1 - p)
Binomial	np	np(1 - p)

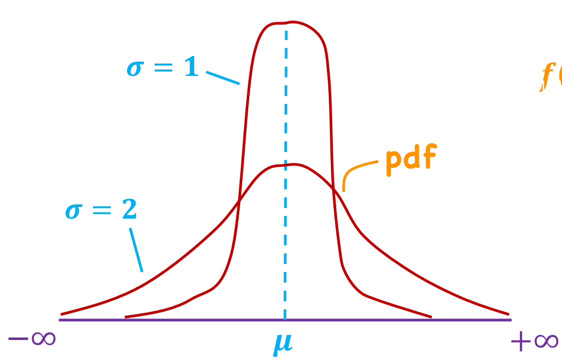
→ p → 1, (1-p) → 0

$$E(x) = p(1) + (1-p)0 = p$$

$$\sigma^2 = p(1-p)^2 + (1-p)(0-p)^2 = p(1-p)(1-p) + (1-p)\,p^2 = p - 2p^2 + p^3 + p^2 - p^3$$
$$= p - p^2 = p(1-p)$$

LOS f
- explain

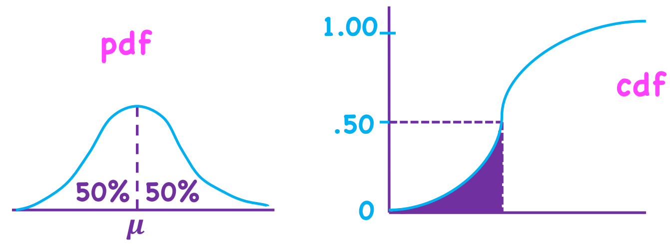
Central Limit Theorem (next reading) → the distribution of a large number of independent random variables with finite variance is approximately normal



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad -\infty < X < \infty$$

- if $\mu = 0, \sigma = 1$: standard normal dist.

Page 9
LOS f
- explain



- the normal dist. will be used to model asset returns (not asset prices)
  - more kurtotic than normal
  - options add skew

- Characteristics:
  1/ described by 2 parameters → μ and σ^2 $\quad X \sim N(\mu, \sigma^2)$
  2/ skew = 0 and kurtosis = 3 $(k_e = 0)$
  ∴ mean = median = mode
  3/ a linear combination of 2 or more normal random variables is also normally distributed

$$\underbrace{R_p}_{\text{normally dist.}} = \underbrace{w_1R_1 + w_2R_2 + w_3R_3}_{\text{univariate random vars.}}$$

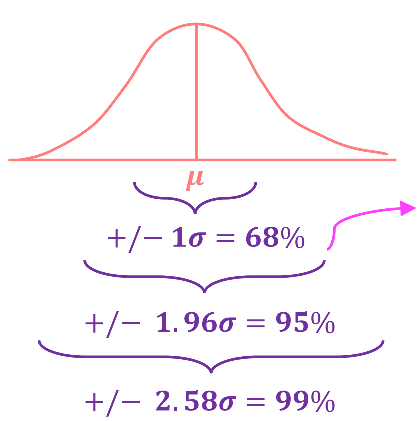
- but multivariate

Page 10
LOS g
- explain
- contrast

- multivariate normal distribution is completely defined by 3 lists of variables:
  1/ all the mean returns of all the individual securities (n returns)
  2/ all the securities' variances (n variances)
  3/ all pairwise correlations $(n^2 - n)/2$ unique correlations

---

LOS h
- calculate



example #5/

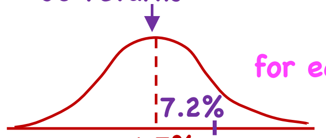
z-value

Ex 6, #1	=NORM.S.DIST(-1,1)	0.158655254
	=1 - NORM.S.DIST(1,1)	
Ex 6, #2	=NORM.S.DIST(1,1) - NORM.S.DIST(-1,1)	
	0.682689492	
Ex 6, #3	=NORM.S.DIST(-2,1)	0.022750132
	=1 - NORM.S.DIST(2,1)	
Ex 6, #4	=NORM.S.INV(0.05)	-1.644853627
	=-NORM.S.INV(0.95)	

1 = cdf
0 = pdf

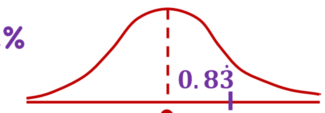
Page 11
LOS i
- explain

n = 30 returns



for each x_i : $z = \dfrac{x_i - \bar{x}}{\sigma}$

e.g./ x = 7.2% $z = \dfrac{7.2 - 4.7}{3} = 0.8\dot{3}$

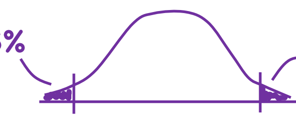


LOS j
- calculate
- interpret

= NORM.S.DIST (z,1)   or   = NORM.S.INV (probability)
(z in, prob. out)   (prob. in, z out)

1/ $p(z \le 0.24)$ = NORM.S.DIST (0.24,1)   0.5984
2/ $p(z \le -1.65)$ = NORM.S.DIST (-1.65,1)   .04947 $\sim 5\%$
3/ 90th percentile = NORM.S.INV (0.90)   1.28155   (z-value)
4/ 95th percentile = NORM.S.INV (0.95)   1.64485

= 1 - NORM.S.INV (0.95)
= NORM.S.INV (0.05)



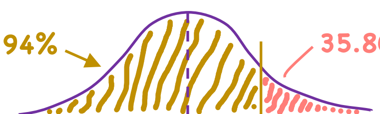
Page 12
LOS j
- calculate
- interpret

Example #6/ $R_P = 12\%$ $\sigma_P = 22\%$

1/ $P(R_P \ge 20\%)$

$z = \dfrac{20 - 12}{22} = 0.3636$

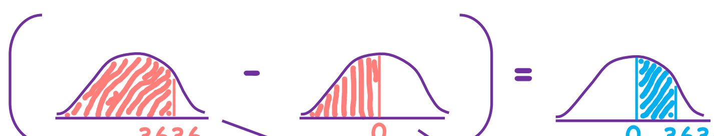
= NORM.S.DIST (0.3636,1)
= 0.64194



$\therefore P(R_P \ge 20\%) = 1 - 0.64194 = .35806$

2/ $P(12\% \le R_P \le 20\%)$

$z = 0$ $z = 0.3636$

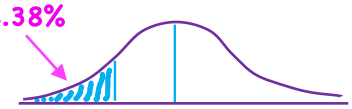


= NORM. S. DIST (0.3636,1) - NORM. S. DIST (0,1)   (= .64194 - .5000 = .14194)

3/ $P(R_P \le 5.5\%)$

$z = \dfrac{5.5 - 12}{22} = -.2955$

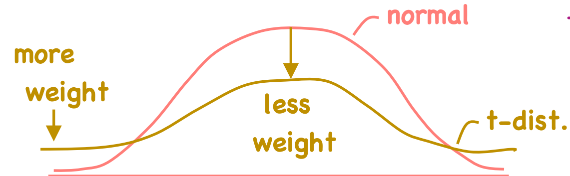
= NORM.S.DIST (-.2955,1)
= .3838



1/ **Student's t-distribution** – defined by a single parameter known as degrees of freedom (df = $n-1$)

LOS k
- describe
- calculate
- interpret

- as $n\uparrow$, the t-distribution approaches the z-distribution
- for $n > 200$, $t \simeq z$



$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{T}} \longrightarrow \text{standard error}$$

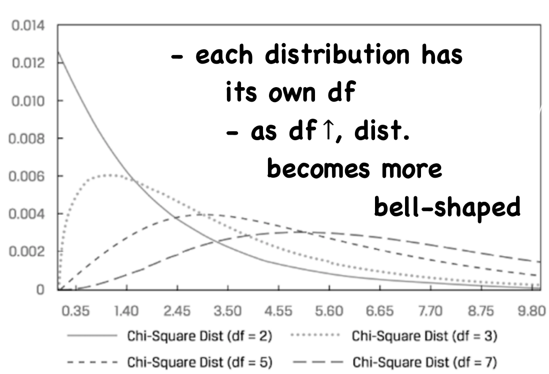
- where μ and σ are population parameters (only 1 estimate used)

$$t = \frac{\overline{X} - \mu}{S/\sqrt{t}}$$

- where $\overline{X}$ and S are sample statistics ($\therefore$ 2 estimates used)

LOS k
- describe
- calculate
- interpret

- t-tests are used for hypothesis testing since they are more conservative, a more stringent test, and they produce wide confidence intervals

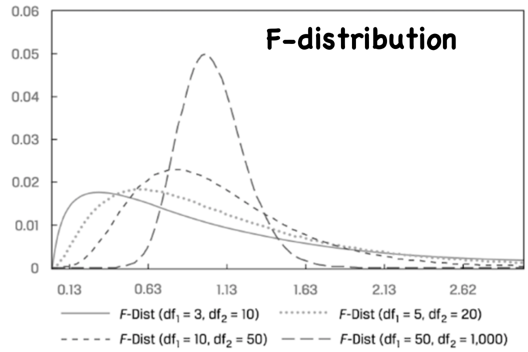


- distribution of the sum of squares (deviations) of k independent standard normally distributed random variables (dist. of variances)

df = $n-1$

- bounded below by zero

LOS L
- describe
- calculate
- interpret



- the ratio of 2 X^2 variables

$$F = \frac{\frac{X_1^2}{n_1 - 1}}{\frac{X_2^2}{n_2 - 1}}$$

→ the larger value is in the numerator

LOS k
- describe
- calculate
- interpret

• used in regression to test the significance of the whole regression (explained var./unexplained var.)

Example 9 Excel cdfs:

NORM.S.DIST (z,1) NORM.S.INV (p)
T.DIST (t-value,df,1) T.INV (p,df)
CHISQ.DIST (x^2-value,df,1) CHISQ.INV (p,df)
F.DIST (F-value,df_1,df_2,1) F.INV (p,df_1,df_2)

Example 9/

LOS k
- describe
- calculate
- interpret

=T.DIST(1,200,1)	0.84074			
=T.DIST(2,200,1)	0.97657			
=T.DIST(3,200,1)	0.99848			
=NORM.S.DIST(1,1)	0.84134	90%	=NORM.S.INV(.9)	1.282
=NORM.S.DIST(2,1)	0.97725	95%	=NORM.S.INV(.95)	1.645
=NORM.S.DIST(3,1)	0.99865	99%	=NORM.S.INV(.99)	2.326
=T.DIST(1,5, 1)	0.81839	90%	=T.INV(.9,5)	1.476
=T.DIST(2,5,1)	0.94903	95%	=T.INV(.95,5)	2.015
=T.DIST(3,5, 1)	0.98495	99%	=T.INV(.99,5)	3.365
=CHISQ.DIST(1,5,1)	0.037	90%	=CHISQ.INV(0.9,5)	9.236
=CHISQ.DIST(2,5,1)	0.151	95%	=CHISQ.INV(0.95,5)	11.070
=CHISQ.DIST(3,5,1)	0.300	99%	=CHISQ.INV(0.99,5)	15.086
=F.DIST(1,5,1,1)	0.363	90%	=F.INV(.9,5,1)	57.240
=F.DIST(2,5,1,1)	0.511	95%	=F.INV(.95,5,1)	230.162
=F.DIST(3,5,1,1)	0.589	99%	=F.INV(.99,5,1)	5763.650

Sampling and Estimation

a. identify and describe desirable properties of an estimator

b. contrast a point estimate and a confidence interval estimate of a population parameter

c. calculate and interpret a confidence interval for a population mean, given a normal distribution with 1) a known population variance, 2) an unknown population variance, or 3) an unknown population variance and a large sample size

d. describe the issues regarding selection of the appropriate sample size, data snooping bias, sample selection bias, survivorship bias, look-ahead bias, and time-period bias

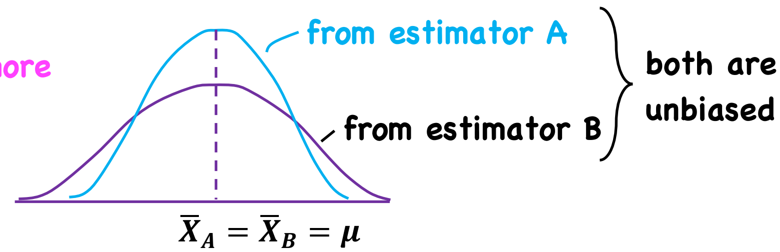
Page 1
LOS a
- identify
- describe

• Point Estimators/ Desirable properties

1/ Unbiasedness → an unbiased estimator is one whose expected value (the mean of its sampling distribution) equals the parameter it is intended to estimate

2/ Efficiency → an unbiased estimator is efficient if no other unbiased estimator has a sampling distribution with smaller variance

• estimator A is more efficient since it produces a smaller variance



Page 2
LOS a
- identify
- describe

• Point Estimators/ Desirable properties

3/ Consistency - a consistent estimator is one for which the probability of estimates close to the value of the population parameter increases as sample size increases

e.g. $SE = S/\sqrt{n}$, as $n \uparrow$, $SE \downarrow$ implying less sampling error in $\bar{X} = \mu$.

• **Example 7**

LOS b, c
- contrast
- calculate
- interpret

Confidence Interval → a range for which one can assert with a given probability $(1-\propto)$, called the degree of confidence, that it will contain the parameter it is intended to estimate

ie. lower limit $\longleftarrow \bar{X} \longrightarrow$ upper limit

• two-sided confidence interval

Page 3
LOS b, c
- contrast
- calculate
- interpret

Confidence Intervals/

Interpretation: Probabilistic → in repeated sampling 95% of such CIs will, in the long run, include or bracket the population mean

Practical → 95% confident that a given CI contains the population mean

$$\underset{\bar{X}}{\text{Point Estimate}} \;\; +/- \;\; \underbrace{\underset{Z_{\propto/2} \text{ or } t_{\propto/2}}{\text{Reliability factor}} \;\; \times \;\; \underset{\sigma/\sqrt{n} \text{ or } S/\sqrt{n}}{\text{Standard Error}}}_{\text{the precision of the estimator}}$$

1/ CI for μ (Normally Distributed population, known variance)

$$\bar{X} \text{ +/- } Z_{\propto/2} \bullet \sigma/\sqrt{n}$$

= NORM.S.INV (.975) = 1.96

or = NORM.S.INV (.025) = -1.96

e.g. $\bar{X} = 25$, $\sigma^2 = 400$, $\propto = 5\%$, $n = 100$

$$25 +/- 1.96(\sqrt{400}/\sqrt{100})$$

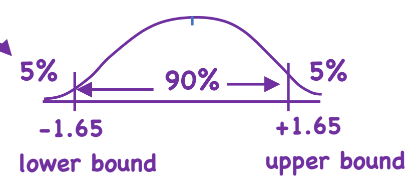
21.08   28.92

Page 4
LOS b, c
- contrast
- calculate
- interpret

1/ CI for μ (Normally Distributed population, known variance)

- common reliability factors

90% = 1.65
95% = 1.96
99% = 2.58



2/ CI for μ (Large sample, Variance Unknown)

- any sampling distribution

$$\bar{X} \text{ +/- } Z_{\propto/2} \;\; S/\sqrt{n}$$

= NORM.S.INV (.95) = 1.65

or = NORM.S.INV (.05) = -1.65

e.g. $\bar{X} = 0.45$ $S = 0.30$ $\propto = 10\%$ $n = 100$

$$.45 +/- 1.65(.3/\sqrt{100})$$

.4005   .4995

Page 5

LOS b, c
- contrast
- calculate
- interpret

3/ CI for μ (σ^2 Unknown)

$\bar{X}$ +/- $t_{\propto/2} \bullet S/\sqrt{n}$

= T. INV (p ,df) = t-val.
or = T. INV (p,df) = -t-val.

• sample is large regardless of distribution
or/ • sample is small but population is normally distributed

	$n < 30$	$n > 30$
normal dist., σ^2 known	z	z
normal dist., σ^2 unknown	t	t or z
non-normal dist., σ^2 known	N/A	z
non-normal dist., σ^2 unknown	N/A	t or z

practice uses t.

• the larger n, the greater the precision

- to obtain a desired CI width, select n as follows:

$\bar{X}$ +/- $\underbrace{t \bullet S/\sqrt{n}}_{E}$

$$E \rightarrow E = \frac{ts}{\sqrt{n}} \rightarrow \sqrt{n} = \frac{ts}{E} \rightarrow n = \left(\frac{t \times S}{E}\right)^2$$

$$n = \left(\frac{t \times S}{E}\right)^2$$

note: width = 2E

Page 6

LOS d
- describe

1/ Data snooping bias - searching a data set for statistically significant patterns/relationships (data mining)

- if $\propto = 5\%$, testing 100 different variables, on average, will produce 5 significant relationships
  - typically will not be theory-driven
  - lack an economic rationale

Page 7
LOS d
- describe

1/ Data snooping bias
- to minimize/avoid:

training data set	validation data set	test data

test data → out-of-sample test to evaluate model fit
- if data snooping is present, there will be insignificant model fit

training data set: build and fit a model

validation data set: evaluate model fit and tune the model

2/ Sample selection bias/ excluding some observations or time periods
- basically choosing non-random samples

e.g./ survivorship bias → historical data may only include data for companies that survived
- would overstate performance

- using hedge fund indexes → since they are self report, only well-performing funds may opt to report

Page 8
LOS d
- describe

3/ Look ahead bias/ using information that was not available on the observation date

e.g.: models that use price and accounting data from the historical record when the actg. data may not have been available on the same date
(P_0 on Dec 31, BV on Dec. 31, but BV may not have been reported until mid-February)

4/ Time-period bias/ results in one time period may be specific to that time period

- typical of short time series
- too long, risk of including more than one regime/distribution

Basics of Hypothesis Testing

a. define a hypothesis, describe the steps of hypothesis testing, and describe and interpret the choice of the null and alternative hypotheses

b. compare and contrast one-tailed and two-tailed tests of hypotheses

c. explain a test statistic, Type I and Type II errors, a significance level, how significance levels are used in hypothesis testing, and the power of a test

d. explain a decision rule and the relation between confidence intervals and hypothesis tests, and determine whether a statistically significant result is also economically meaningful

e. explain and interpret the *p*-value as it relates to hypothesis testing

f. describe how to interpret the significance of a test in the context of multiple tests

g. identify the appropriate test statistic and interpret the results for a hypothesis test concerning the population mean of both large and small samples when the population is normally or approximately normally distributed and the variance is (1) known or (2) unknown

h. identify the appropriate test statistic and interpret the results for a hypothesis test concerning the equality of the population means of two at least approximately normally distributed populations based on independent random samples with equal assumed variances

i. identify the appropriate test statistic and interpret the results for a hypothesis test concerning the mean difference of two normally distributed populations

j. identify the appropriate test statistic and interpret the results for a hypothesis test concerning (1) the variance of a normally distributed population and (2) the equality of the variances of two normally distributed populations based on two independent random samples

Hypothesis Testing

Page 1
LOS a
- define
- describe
- interpret

Statistical Inference → the process of making judgments about a larger group (pop.) based on a smaller group (sample)

e.g./ hypothesis testing - test to see whether a sample statistic is likely to come from a population with the hypothesized value of the population parameter

i.e. Does $\overline{X} = \mu_0$?

Hypothesis → a statement about one or more populations that are tested using sample statistics

Process: Step 1: State the hypothesis
2: Identify the appropriate test statistic
3: Specify the level of significance
4: State the decision rule
5: Collect data and calculate the test statistic
6: Make a decision

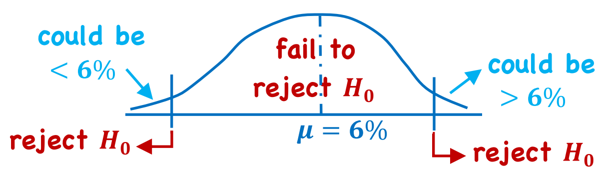
Page 2
LOS a
- define
- describe
- interpret

Step #1: State the hypothesis

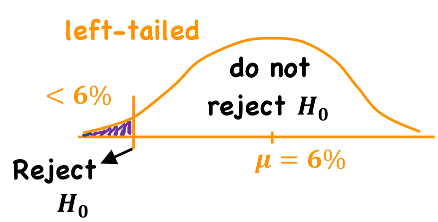
null → H_0 → assumed to be true unless
alternative → H_a we can reject
- typically want to reject H_0

LOS b
- compare
- contrast

- Two-sided (two-tailed) test
  e.g. $H_0 : \mu = 6\%$
  vs. $H_a : \mu \neq 6\%$



- One-sided (left or right tailed) test
  e.g. $H_0 : \mu \leq 6\%$ or/ $H_0 : \mu \geq 6\%$
  $H_a : \mu > 6\%$ $H_a : \mu < 6\%$
  → right-tailed ← left-tailed



Page 3
LOS b
- compare
- contrast

- the null (H_0) always contains the equality sign

$$H_0 : \bar{X} = \mu_0 \qquad H_0 : \bar{X} \le \mu_0 \qquad H_0 : \bar{X} \ge \mu_0$$

- testing H_0 is always done at equality

example #1

Test Statistic:
(Step #2)

pop. σ^2 is known

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

distributed normally

pop. σ^2 is unknown

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

t-distributed

What We Want to Test	Test Statistic	Probability Distribution of the Statistic	Degrees of Freedom
Test of a single mean	$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$	t-Distributed	$n - 1$
Test of the difference in means	$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$	t-Distributed	$n_1 + n_2 - 2$
Test of the mean of differences	$t = \frac{\bar{d} - \mu_{d0}}{s_{\bar{d}}}$	t-Distributed	$n - 1$
Test of a single variance	$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$	Chi-square distributed	$n - 1$
Test of the difference in variances	$F = \frac{s_1^2}{s_2^2}$	F-distributed	$n_1 - 1, n_2 - 1$
Test of a correlation	$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	t-Distributed	$n - 2$
Test of independence (categorical data)	$\chi^2 = \sum_{i=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$	Chi-square distributed	$(r - 1)(c - 1)$

LOS c
- explain

Page 4
LOS c
- explain

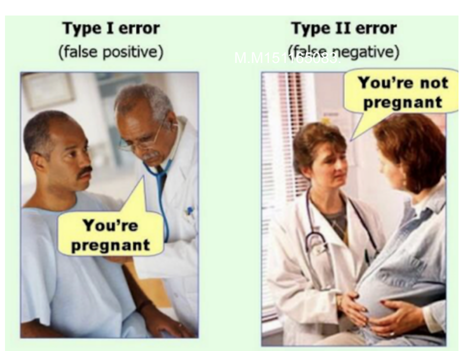
Step 3: Specify the Level of Significance

- level of sig. depends on the seriousness of making a mistake

	H_0 = true	H_0 = false
fail to reject	Correct $(1 - \alpha)$ confidence level	Type II error β
reject	Type I error α level of sig.	Correct $(1 - \beta)$ Power of a test

- as $\alpha \downarrow$, $\beta \uparrow$
- only way to decrease both is to increase n

e.g.: H_0: not pregnant
H_a: pregnant



$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

→ as $n \uparrow$, denom. $\downarrow$, t-stat $\uparrow$

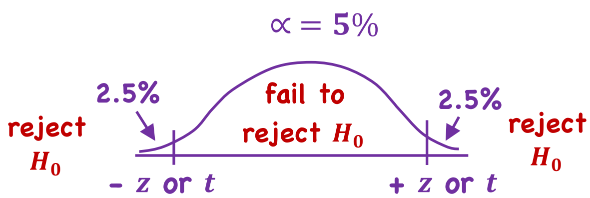
Page 5
LOS d
- explain
- determine

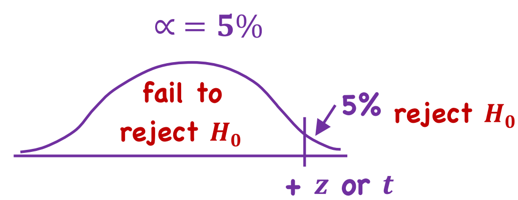
Step #4: State the Decision Rule

2-tail: Reject H_0 when $|\text{test} - \text{statistic}| > |\text{critical value}|$

t or z $\quad$ $t_{\propto/2}$ or $z_{\propto/2}$

right tail: Reject H_0 when test-statistic $>$ critical value
left tail: Reject H_0 when test-statistic $<$ critical value $\quad$ ($t_{\propto}$ or $z_{\propto}$)



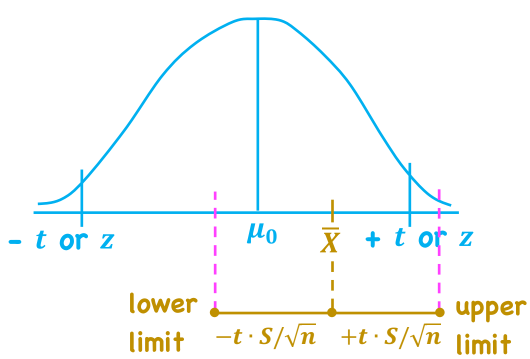


Cut-off for . . .	Excel	Python	R
Right tail, 2.5%	NORM.S.INV(0.975)	norm.ppf(.975)	qnorm(.025,lower.tail=FALSE)
Left tail, 2.5%	NORM.S.INV(0.025)	norm.ppf(.025)	qnorm(.025,lower.tail=TRUE)
Right tail, 5%	NORM.S.INV(0.95)	norm.ppf(.95)	qnorm(.05,lower.tail=FALSE)
Left tail, 5%	NORM.S.INV(0.05)	norm.ppf(.05)	qnorm(.05,lower.tail=TRUE)

or T.INV (p,df)

Page 6
LOS d
- explain
- determine

- using confidence intervals: $\bar{X} +/- t \bullet S/\sqrt{n}$



- if the CI around the sample statistic ($\bar{X}$) contains the hypothesized pop. parameter,
  Do not reject H_0

Step 5: Collect the data and Calculate the test statistic $\quad$ Ex #3

Step 6: Make a decision

- Reject H_0 if there is statistical support

Note: something statistically significant may not be economically significant (transaction costs, taxes, risk)

ex #4

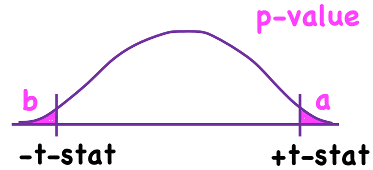
Page 7
LOS e
- explain
- interpret

P-value → the area in the probability distribution outside the calculated test-statistic

- for a two-sided test-stat, combine the probabilities under the curve in both tails

= (1 - NORM.S.DIST(+z,1)) x 2
or/ = (1 - T.DIST(+t,df,1) x 2

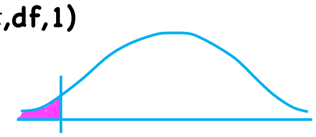
p-value = a + b



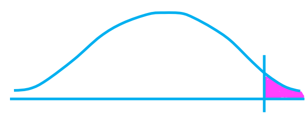
- p-value is the lowest ∝ at which the null can be rejected

- one-sided: prob. under the curve in the appropriate tail

= NORM.S.DIST(-z,1)
= T.DIST(-t,df,1)



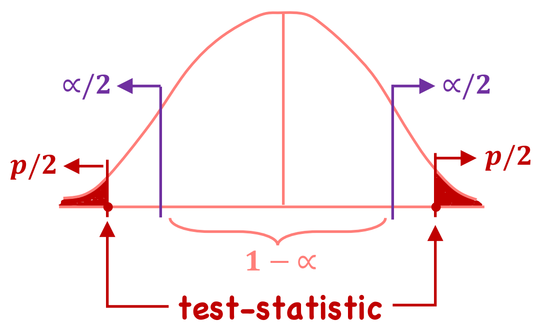
= 1 - NORM.S.DIST(+z,1)
= 1 - T.DIST(+t,df,1)



Page 8
LOS e
- explain
- interpret

- if p-value < ∝ , reject H_0

$$2\left(\frac{p}{2}\right) < 2\left(\frac{\propto}{2}\right)$$
$$= p < \propto$$



example #5

LOS f
- describe

- Rejecting H_0 is a positive event (support for H_a)
- Rejecting a true H_0 is thus a false positive

- FDR - false discovery rate → expected proportion of false positives
- multiple testing problem → in m repeated tests with a level of significance ∝, expect to generate $m \propto$ false positives
  - this requires an adjustment to the p-value for the likelihood of significant results being false positives (BH - Benjamini & Hochberg)

Page 9
LOS f
- describe

- rank all p-values from low to high
  - starting at the lowest:

Let $k = 1$, check $p(1) \leq \propto \frac{\text{rank of } i}{\text{\# of tests}}$

yes, Let $k = k + 1$, repeat
no, stop, Let $k = k$

k = adjusted # of FPs.

exhibit #12 + example #6

LOS g
- identify
- interpret

Tests concerning a single mean/

		< 30	≥ 30
population	σ^2 known	t	z
approx. normal	$\underbrace{\sigma^2 \text{ unknown}}_{\text{this is typically the case}}$	t	t or z

→ recall CLT-sampling distribution of means will be approx. normal with large n regardless of pop. dist.

$\therefore$ test statistic $= t_{n-1} = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$

Page 10
LOS g
- identify
- interpret

z test-stat. σ^2 known
- theoretically correct to use:

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

z-alternative: σ^2 unknown
but large n:

$$z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Decision Rule: if $|\text{test-stat}| > |\text{critical value}|$ → reject H_0
2-sided

= NORM. S. DIST(p)
= T. DIST(p,df)

$\left.\begin{array}{l}\text{test-stat} > \text{critical value} \text{ - right}\\ \text{test-stat} < \text{critical value} \text{ - left}\end{array}\right\}$ reject R_0

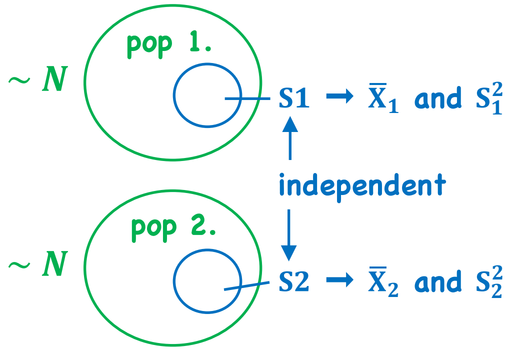
- common z-values

1%	2-tailed	+/- 2.576
	1-tailed	-2.326 or +2.326
5%	2-tailed	+/- 1.96
	1-tailed	-1.645 or +1.645

example #7
example #8

Page 11
LOS h
- identify
- interpret

• **Differences between means - independent samples**



Q: Are $\overline{X}_1$ and $\overline{X}_2$ from the same population (i.e. $\overline{X}_1 = \overline{X}_2$) or from different populations (i.e. $\overline{X}_1 \neq \overline{X}_2$)

Assumption: $S_1^2 = S_2^2$

test statistic:

$$t_{n_1+n_2-2} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Ex #9

2-sided

$H_0: \mu_1 - \mu_2 = 0$
$H_0: \mu_1 - \mu_2 \neq 0$

or/

$H_0: \mu_1 = \mu_2$
$H_a: \mu_1 \neq \mu_2$

1 sided - right

$H_0: \mu_1 - \mu_2 \leq 0$
$H_a: \mu_1 - \mu_2 > 0$

$H_0: \mu_1 \leq \mu_2$
$H_a: \mu_1 > \mu_2$

1-sided - left

$H_0: \mu_1 - \mu_2 \geq 0$
$H_a: \mu_1 - \mu_2 < 0$

$H_0: \mu_1 \geq \mu_2$
$H_a: \mu_1 < \mu_2$

Page 12
LOS i
- identify
- interpret

• **Differences between means - dependent samples**

• have something in common
• equal n (i.e. $n_1 = n_2$)
- paired observations

- arrange data in pairs
- calculate a new variable

$$d = X_{A_i} - X_{B_i}$$

and $\overline{d} = \frac{\sum d_i}{n}$ and $SE = S_d/\sqrt{n}$ n = # of paired observations

$H_0: \mu_d = \mu_{d_0}$
$H_a: \mu_d \neq \mu_{d_0}$

$H_0: \mu_d \leq \mu_{d_0}$
$H_a: \mu_d > \mu_{d_0}$

$H_0: \mu_d \geq \mu_{d_0}$
$H_a: \mu_d < \mu_{d_0}$

test-statistic: $t_{n-1} = \frac{\overline{d} - \mu_{d_0}}{S_{\overline{d}}/\sqrt{n}}$

Example 10, 11

Page 13
LOS j
- identify
- interpret

- **Tests of Variances**

1/ Single Variance - n independent observations from a normally distributed pop.
- chi-square tests sensitive to violations

- not symmetrical, ∴ critical values are also not symmetrical

= CHISQ.INV(lower p,df) → lower X^2
= CHISQ.INV(upper p,df) → upper X^2

2 sided:
$H_0: \sigma^2 = \sigma_0^2$
$H_a: \sigma^2 \neq \sigma_0^2$

1 sided: left
$H_0: \sigma^2 \geq \sigma_0^2$
$H_a: \sigma^2 < \sigma_0^2$

right
$H_0: \sigma^2 \leq \sigma_0^2$
$H_a: \sigma^2 > \sigma_0^2$

test-statistic = $X_{n-1}^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

example #12

Page 14
LOS j
- identify
- interpret

- **Tests of Variances** (e.g. compare volatility of 2 funds)

2/ Equity of 2 Variances

$H_0: \sigma_1^2 = \sigma_2^2$
$H_a: \sigma_1^2 \neq \sigma_2^2$

$H_0: \sigma_1^2 \leq \sigma_2^2$
$H_a: \sigma_1^2 > \sigma_2^2$

$H_0: \sigma_1^2 \geq \sigma_2^2$
$H_a: \sigma_1^2 < \sigma_2^2$

or/ $H_0: \sigma_1^2/\sigma_2^2 = 1$
$H_a: \sigma_1^2/\sigma_2^2 \neq 1$

$\underbrace{H_0: \sigma_1^2/\sigma_2^2 \leq 1 \quad H_a: \sigma_1^2/\sigma_2^2 > 1}_{\text{right-tailed}}$

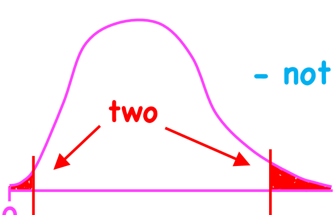
$\underbrace{H_0: \sigma_1^2/\sigma_2^2 \geq 1 \quad H_a: \sigma_1^2/\sigma_2^2 < 1}_{\text{left-tailed}}$

test-statistic

$F = S_1^2/S_2^2$

$\text{df}_1 = n_1 - 1$
$\text{df}_2 = n_2 - 1$

- with the larger S^2 in the numerator



- not symmetrical
∴ need 2 critical values

= F. INV(lower p,df_1,df_2)
= F. INV(upper p,df_1,df_2)

Example 13/14