Ch9: Decision Making
Ch1-6 Review

- *REB due today*
- *No HW this week!*
- *Midterm ch1-6 next Tue in-class*

14 Feb 2012
Dr. Sean Ho

busi275.seanho.com

Trinity Western University

Outline for today

- Decision making
  - Null hypothesis (H_0) vs. alternate (H_A)
  - "Reject H_0" vs. "Fail to reject H_0"
  - Risks of error: Type I and Type II error
- Hypothesis tests
  - On population mean (μ), with known σ
  - On μ, with unknown σ (TDIST)
  - On binomial proportion π
- Midterm review: ch1-6

Decision making

- The real world is fuzzy / uncertain / complex
- To make decisions, we need to assess risk
  - Fuzzy risk → binary yes/no decision
- A hypothesis is an idea of how the world works
  - Decision: accept or reject the hypothesis?
  - Based on the data, what are the risks in accepting hypothesis? Risks in rejecting?
- Null hypothesis (H_0) is the default, “status quo”
  - Fallback if insufficient evidence for H_A
- Alternate hypothesis (H_A) is the opposite
  - Usually same as our research hypothesis: what we intend to show

H_0 vs. H_A

- Do actively-managed mutual funds outperform the market (measured by index funds)?
  - H_0: no difference, or do not outperform
  - H_A: do outperform
- Does gender affect investment risk tolerance?
  - H_0: no difference, tolerance same for both
  - H_A: risk tolerance of men + women differs
- Supplier claims defect rate is less than 0.001%
  - H_0: defect rate is too high: $\geq$ 0.001%
  - H_A: supplier has proved defect rate is low

"Reject H_0" vs. "fail to rej H_0"

- Two options for making decisions:
  - Reject H_0: strong statement, significant evidence in favour of H_A and against H_0
  - Fail to reject H_0: weak statement, insufficient evidence in favour of H_A
    - Does not mean strong evidence in favor of accepting H_0! Perhaps need more data
- Mutual funds: "reject H_0" means strong evidence that active management beats market
  - "Fail to reject H_0" means insufficient evidence to show they perform better

Risks / errors

- Our decision may or may not be correct:

	H_0 true	H_A true
Rej H_0	Type I	
Fail rej H_0		Type II

- We define H_0/H_A so that Type I error is worse and Type II error is more bearable
  - Can't eliminate risk, but can manage it
  - α is our limit on Type I (level of significance)
  - β is our limit on Type II ($1-\beta$ = "power")

Type I vs. Type II risks

- Supplier: H_0: high defect rate; H_A: low defects
  - Type I: think defect rate is low, when in reality it is high: ⇒ angry customers
  - Type II: supplier is good, but we wrongly suspected / fired them: ⇒ loss of partner
- Murder trial: H_0 / H_A? Type I/II?
- Parachute inspector: H_0 / H_A? Type I/II?
- In most research, $\alpha=0.05$ and β is unlimited
  - But depends on context, meaning of H_0/H_A
  - e.g., what should α for parachute be?

Outline for today

- Decision making
  - Null hypothesis (H_0) vs. alternate (H_A)
  - "Reject H_0" vs. "Fail to reject H_0"
  - Risks of error: Type I and Type II error
- Hypothesis tests
  - On population mean (μ), with known σ
  - On μ, with unknown σ (TDIST)
  - On binomial proportion π
- Midterm review: ch1-6

Steps in decision making

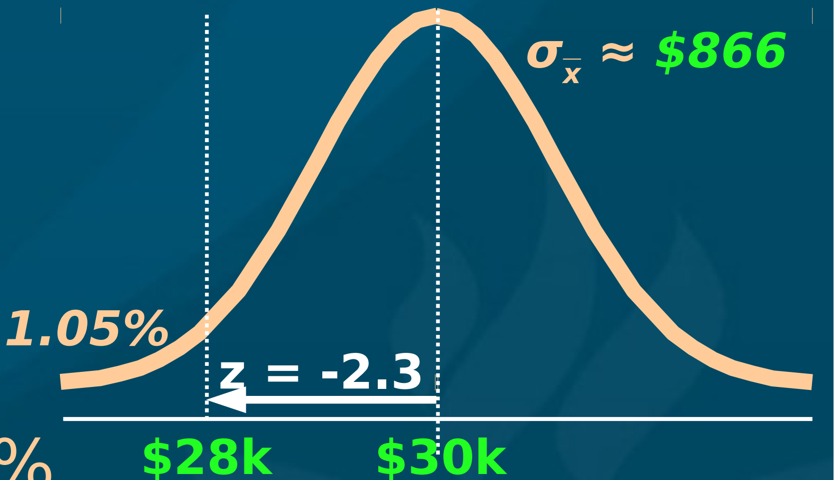
1) Research question: variables, levels of meas
2) H_0 vs. H_A: in words and notation
    1) What kind of hypothesis test is needed?
    2) What does it assume?
3) Apply hypothesis test:
    1) Find standard error if appropriate
    2) Find chance of Type-I error (*p*-value)
4) State conclusion and interpret in context:
    - If $p < \alpha$, then reject H_0
    - Otherwise, fail to reject H_0

Example: starting salary

- Assume starting salary of clerical workers is normally distributed with σ=\$3k.
  Question: is their avg salary significantly less than \$30k?
  - Data: sample $n=12$ salaries, get $\bar{x}$=\$28k
- Only one variable: salary (ratio)
  - Making an inference about its mean: μ
- State hypotheses:
  - H_0 (default fallback): avg salary $\mu \geq \$30k$
  - H_A (research hypothesis): $\mu < \$30k$

Salary example: *p*-value

- Calculate risk of Type I error (*p*-value)
  - Assume μ is what H_0 says it is (μ=$30k)
  - Sample data $\bar{x}$ is a threshold on the SDSM
  - Risk of Type I error is area in tail of SDSM
- Standard error (SE): $\sigma_{\bar{x}}$ = σ/√n = $3k/√12 ≈ $866
- Z-score: (28-30)/866 ≈ 2.3
- *p*-value is the area in tail:
  - NORMSDIST(-2.3) → 1.07%
- Or, doing this all in one calculation:
  - NORMDIST(28, 30, 3/SQRT(12)) → 1.05%



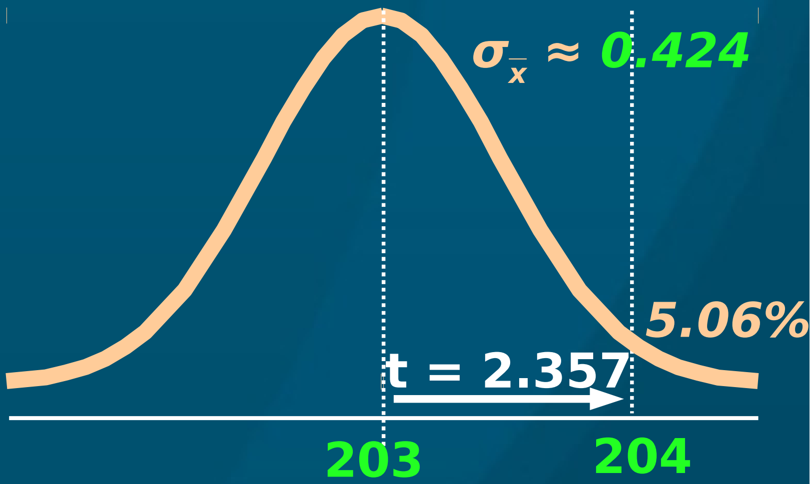
Salary example: conclusion

- So, if the average salary were truly $\geq$ \$30k, then there would only be a p=1.05% chance that we might randomly pick a sample that has $\bar{x}$ at or lower than \$28k
  - Conclude that true mean is probably <\$30k
- We use this 1.05% as our risk of Type I error:
  - Compare against α (usually 5%)
  - Conclude this is an acceptable risk, so
- Reject H_0: at the 5% level of significance, the starting salaries of these clerical workers are significantly lower than \$30k

Two-tailed tests

- The preceding example was "one-tailed"
    - H_0 / H_A use directional inequalities $<, \leq, >, \geq$
    - "greater than", "bigger", "more/less"
- Two-tailed test uses non-directional inequalities
    - $\neq$, "differ", "change", "same / not same"
- e.g., standard height of doors is 203cm.
  Is a batch of doors significantly out of spec?
    - H_0: no difference, within spec: $\mu = 203$ cm
    - H_A: either too tall or too short: $\mu \neq 203$ cm
    - Data: measure a sample of doors, get n, $\bar{x}$, s
        - Say $n=8$, $\bar{x}=204$, $s=1.2$

Door ex.: two-tailed, no σ



- Std err = $s/\sqrt{n} \approx 0.424$
- t = (204-203)/0.424 ≈ 2.357
- Apply TDIST with df=7 to find the % in both tails:
  - TDIST(2.357, 7, 2) → 0.0506
  - More precisely:
    TDIST((204-203) / (1.2/SQRT(8)), 7, 2)
- So our calculated risk of Type I error is 5.06%
  - Assuming normal distribution of door height
- This is larger than our tolerance (α):
  - Unacceptably high risk of Type I error

Door ex.: conclusion

- In view of the high risk of Type I error, we are unwilling to take that risk, so we conclude:
  - Fail to reject H_0: at the 5% level, this batch of doors is not significantly out of spec
- In this example, we follow the research convention of assigning '=' to H_0
  - But in quality control (looking for defects), we might want H_0 to assume there is a defect, unless proven otherwise
- Also note that if this test had been one-tailed:
  - TDIST(2.357, 7, 1) → 2.53% < α and we would have rejected H_0!

Test on binomial proportion π

- p.373 #33: Wall Street Journal claims 39% of consumer scam complaints are on identity theft
  - RQ: do we believe the claim? H_A: $\pi \neq 0.39$
  - Data: 40/90 complaints are about ID theft
- Std err: $\sigma_p = \sqrt{(pq/n)} = \sqrt{(.39*.61/90)} \approx .0514$
- Z-score: $z = (40/90 – .39) / .0514 \approx 1.06$
- P-value (two-tailed): 2*NORMSDIST(-1.06)
  - Or: 2*(1-NORMDIST(40/90, .39, .0514, 1))
  - → 28.96%
- Fail to reject H_0: insufficient evidence to disbelieve the WSJ claim, so we believe it

Outline for today

- Decision making
  - Null hypothesis (H_0) vs. alternate (H_A)
  - “Reject H_0” vs. “Fail to reject H_0”
  - Risks of error: Type I and Type II error
- Hypothesis tests
  - On population mean (μ), with known σ
  - On μ, with unknown σ (TDIST)
  - On binomial proportion π
- Midterm review: ch1-6

Midterm review: ch1-6

- Ch1: Basic vocabulary: variables and sampling
- Exploring your data:
  - Ch2: using charts
  - Ch3: using descriptives
- Ch4: Probability and independence
- Common distributions:
  - Ch5: Discrete: binom, Poisson, hypgeom
  - Ch6: Continuous: norm, unif, expon

Ch1: Introduction

- Population vs. sample
  - Sampling, inference
  - Statistics, parameters
- Sampling
  - Kinds of bias in collecting data
- 4 levels of measurement

Ch2-3: Exploring Data

- For nominal variables:
  - Charts: bar/col, pie
    - Joint distrib of 2 vars: pivot table
  - Stats: frequency distribution
- For quantitative (interval/ratio/scale) vars:
  - Charts: histogram, ogive (cumul), boxplot
    - Joint distrib of 2 vars: scatter
    - Time series: line
  - Centre: mean, median, mode, (skew)
  - Quantile: Q_1/Q_3, %-ile, IQR
  - Std dev: σ, s, CV, empirical rule, z-score

Ch4: Probability

- Tree diagrams
- P(A) notation, Venn diagrams
  - Sample space, outcome, event
  - $\cap$, $\cup$, complement
- Addition rule: $A \cup B = A + B - (A \cap B)$
  - Mutual exclusivity
- Conditional probability
  - What does it mean; how to find it (Bayes)
  - Statistical independence
    - Does $P(A|B) = P(A)$?

Ch5: Discrete distributions

- Binomial: BINOMDIST(x, n, p, cum)
  - x: counting # of successes out of n trials
  - p: probability of success (binom proportion)
- Poisson: POISSON(x, λ, cum)
  - x: # occurrences within the time period
  - λ: mean (expected) # occ w/in the period
- Hypergeometric: HYPGEOMDIST(x, n, X, N)
  - x, n: # successes & tot size of sample
  - X, N: # successes & tot size of population
    - Binomial p = X/N

Ch6: Continuous distributions

- Normal: NORMDIST(x, μ, σ, cum)
  - Also NORMINV(area, μ, σ), NORMSDIST(z), NORMSINV(area)
- Uniform:
  - $P(x) = 1/(b-a)$, $\mu = (a+b)/2$, $\sigma = \sqrt{(b-a)^2/12}$
- Exponential: EXPONDIST(x, λ, cum)
  - x: time between occurrences
  - λ: 1 / (mean time between occurrences)
    - λ = expected frequency of occurrences (e.g., occurrences per min)

TODO

- REB application due today
  - If not REB exempt, need printed signed copy
- Midterm ch1-6 next classtime