Decision Making (Hypothesis Testing)

- **HW5 due Thu**
- **Work on REB forms**

11 Oct 2011
BUSI275
Dr. Sean Ho

TRINITY WESTERN UNIVERSITY

Outline for today

- Decision making and hypothesis testing
  - Null hypothesis (H_0) vs. alternate (H_A)
- Making conclusions:
  - "reject H_0" vs. "fail to reject H_0"
- Risks of error: Type I and Type II error
- Hypothesis test on population mean (μ)
  - One-tailed vs. two-tailed
- Test on μ, with unknown σ (TDIST)
- Test on binomial proportion π

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 index funds outperform actively-managed mutual 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
- Index funds: “reject H_0” means strong evidence that index outperforms active management
  - “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	✓	✗ 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")

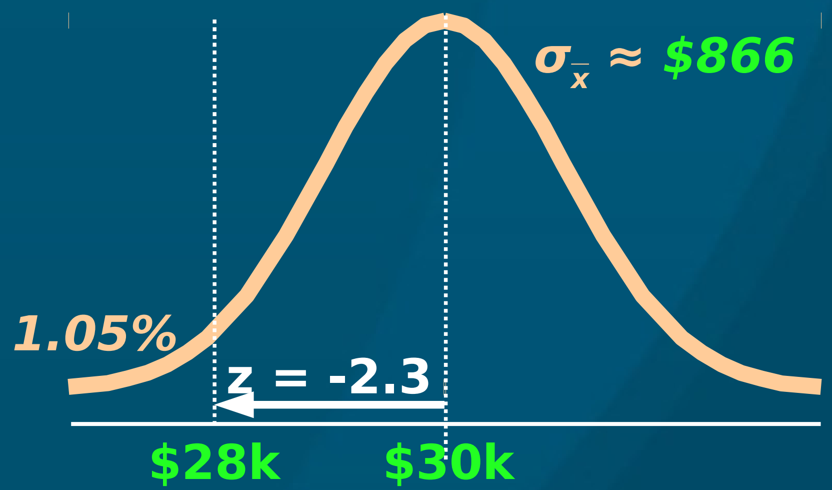
Type I vs. Type II risks

- 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?

Test on population mean

- e.g., assume starting salary of clerical workers is normally distributed with σ=\$3k
  - Research question: is avg salary < \$30k?
  - Data: sample $n=12$ salaries, get $\bar{x}$=\$28k
- H_0 (status quo): avg salary $\mu \geq$ \$30k
  - H_A (research hypothesis): $\mu <$ \$30k
- Strategy: 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

Test on pop mean, cont.

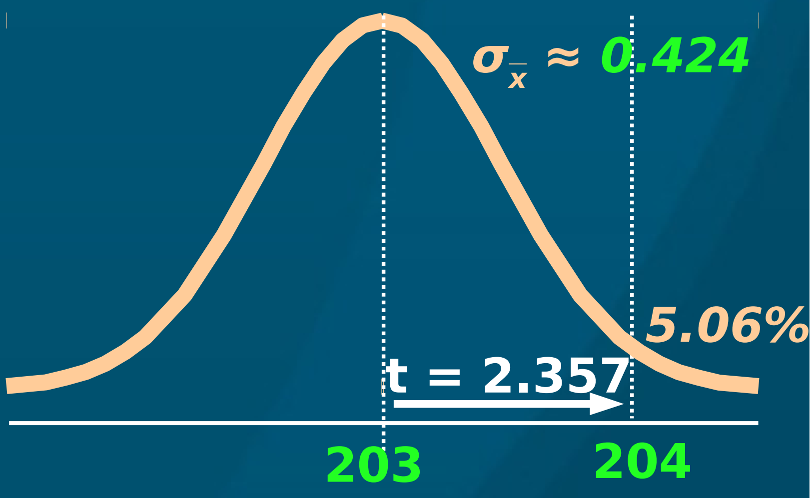


- In our example,
- Std err $\sigma_{\bar{x}} = \sigma/\sqrt{n}$
  $= \$3k/\sqrt{12} \approx \$866$
- Z-score: $(28-30)/866 \approx 2.3$
- Area in tail: NORMSDIST(-2.3) → 1.07%
  - Or: NORMDIST(28, 30, 3/SQRT(12)) → 1.05%
- So there is a 1.05% risk of Type I error
  - Compare against α (usually 5%)
  - Conclude this is an acceptable risk, so
- Reject H_0: yes, at the 5% level of significance, salaries 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: differ from spec
    (either too tall or too short): $\mu \neq 203$ cm
  - Data: measure a sample of doors, get n, $\bar{x}$, s

Door ex.: two-tailed, no σ



- H_A: $\mu \neq 203$
- Data: $n=8$, $\bar{x}=204$, $s=1.2$
- Std err = $s/\sqrt{n} \approx 0.424$
- $t = (204-203)/0.424 \approx 2.357$
- df=7, so the % in both tails (*p*-value) is
  - TDIST(2.357, 7, 2) → 0.0506
  - More precisely: TDIST(1/(1.2/SQRT(8)), 7, 2)
- So our calculated risk of Type I error is 5.06%
  - Assuming normal distribution
- 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 chance, 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_A
  - 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

TODO

- HW5 (ch7-8): due this Thu at 10pm
- REB form due next Tue 18 Oct 10pm
  - If approval by TWU's REB is required, also submit printed signed copy to me
  - You are encouraged to submit early to allow time for processing by TWU's REB (3-4 weeks)
- Midterm (ch1-8): next week Thu