Financial Econometrics | Lecture 2

Review of Statistics: Estimation, Testing and Confidence Intervals

In Class Group Activity

· SJTU, Summer School 2026 · Fatih Kansoy

Lesson of the day. An estimate is one number with a margin of error around it, and a hypothesis test asks one disciplined question: could chance alone have produced a gap this big? So "significant" means "unlikely to be luck", not "large" and not "true".

Phase 1 – Concept checks

ANSWERS

ConcepTest 1. **Answer.** Option **(c)**. The p-value lives in a world where the null is true; it measures the data, not the hypothesis. Option (a) reads it as the probability the null is true; (b) as the error rate of the whole decision; (d) as the probability the alternative is true. All three attach a probability to a hypothesis, which the p-value never does.

A small p means: this data would be a surprise if nothing were going on.

ConcepTest 2. **Answer.** Option **(c)**. Significance answers "is it likely real?"; importance answers "is it big enough to care about?". With 50,000 trades even a useless 0.04% becomes significant (option a and d confuse a big t or a big n with a big effect); option (b) treats "not significant" as "no effect", when 12 trades is simply too few to tell. Keep the two questions apart.

ConcepTest 3. **Answer.** Option **(c)**. The true mean is a fixed number and has no probability; the randomness is in the interval, which jumps around sample to sample. The 95% is the long-run hit rate of the method. Option (a) attaches a probability to this one fixed interval; (b) confuses a confidence interval for the mean with the spread of the raw data; (d) is trivially wrong, since the sample mean is the known centre.

Phase 2 – The desk problem

WORKED SOLUTIONS

(a) $SE = \dfrac{1.50}{\sqrt{10}} = \dfrac{1.50}{3.1623} = 0.4743$. The 95% interval is $1.20 \pm 1.96 \times 0.4743 = 1.20 \pm 0.930 =$ [0.27%, 2.13%]. The half-width 0.93 is the margin of error, which is just 1.96 times the wobble you met in Lecture 1.

(b) $t = \dfrac{1.20 - 0}{0.4743} = 2.530$. Since $2.530 > 1.96$, **reject** H_0 at 5%. The two-sided p-value is about $2 \times [1 - \Phi(2.53)] = 2 \times 0.0057 \approx 0.011$. If the true mean were zero, a gap this big would happen about 1 time in 90 by luck. Zero sits *outside* the interval $[0.27, 2.13]$, which is the same verdict by construction: the confidence interval is the set of nulls you cannot reject.

(c) $SE_{\text{diff}} = \sqrt{\dfrac{3.0^2}{36} + \dfrac{2.4^2}{36}} = \sqrt{0.25 + 0.16} = \sqrt{0.41} = 0.6403$. The difference is $1.50 - 0.60 = 0.90$, so $t = \dfrac{0.90}{0.6403} = 1.406$. Since $|t| = 1.41 < 1.96$, **fail to reject**; the p-value is about 0.16.

Helix did beat Pinnacle by 0.9% a month in the sample, but with this much noise a gap that size

turns up about 1 *time in* 6 *even when the desks are truly equal. "Fail to reject" does not mean the desks are equal; it means 36 months is not enough to tell them apart.*

(d) (1) With $t_9 = 2.262$, the interval is $1.20 \pm 2.262 \times 0.4743 = 1.20 \pm 1.073 = [0.13\%, 2.27\%]$, *wider* than (a). (2) The t-statistic 2.530 still exceeds 2.262, so you **still reject**, but only just. (3) With $\overline{Y} = 1.05$ and the same SE, $t = 2.214$: the lazy 1.96 rule says **reject** ($2.214 > 1.96$), but the correct $t_9 = 2.262$ rule says **fail to reject** ($2.214 < 2.262$). Same data, opposite verdict, because with only ten months the extra uncertainty in s raises the bar. As n grows, $t_{n-1} \to 1.96$.

Phase 3 – The case

WORKED ANSWERS

1. Yes, $t = 2.24 > 1.96$, so it is significant at 5%, but only just.

2. +0.95% a month is about +12% a year, large on paper. Before believing it you would want the trading cost (turnover and capacity): momentum signals decay fast and cost a lot to trade, so the after-cost number can be far smaller.

3. $1 - 0.95^{20} = 1 - 0.358 = \mathbf{0.64}$. With 20 useless signals there is about a 64% chance that at least one looks significant by luck. Reporting only the winner makes its p-value close to meaningless.

4. The pension fund's analyst asked one question: show me 2020 to 2023, out of sample, with the signal fixed in advance. On those next 48 months the same signal earned +0.06% per month, so $SE = 3.6/\sqrt{48} = 0.520$ and $t = 0.06/0.520 = 0.12$, $p \approx 0.90$. The edge had vanished. The in-sample "significance" was an artefact of picking the best of 20 tries on one stretch of data; the true error rate was about 64%, not 5%. The fund did not allocate. Three lessons: significant is not the same as real when you have searched over many candidates; the only honest test is out of sample, fixed in advance; and a barely significant $t = 2.24$ is fragile, because an honest correction for multiple testing pushes it back below the line. This is the mirror image of the medical "absence of evidence": here "reject" did not mean a real effect, because the test was run 20 times and only the winner was shown.

End of worked solutions.