Tapping test (single observation)

- tapping test from Chapter 8
- tapping rate can be used as an indicator for neurological damage
- Christanson & Leathem (2004)
  - 10 s tapping test in normal pop: mean = 59; sd = 7
  - tapping rate is slower in Alzheimer’s patients (& other clinical populations)
  - measured rate for 1 subject = 45 (per 10 s trial)
- is this subject drawn from typical or Alzheimer population?

General Strategy

reject H0 if z exceeds critical values of z

2-tailed test

1-tailed test

Critical values of z

Critical value
Tapping test (single observation)

- measured rate for 1 case = 45 (per 10 s trial)
- is this case unusual?
  - H₀: μ ≥ 59 (case was not drawn from Alzheimer’s population)
  - H₁: μ < 59 (case was drawn Alzheimer’s population)
- significance level of .05 (one-tailed)
  - critical value of z = -1.645; p(z ≤ -1.645 | H₀) = .05
- observed z = (45-59)/7 = -2.0
- observed z < critical z; p(z ≤ -2 | H₀) = 0.023
  - reject H₀ in favour of H₁

Tapping test (group mean)

- Instead of single case study, consider a situation in which we measure the mean tapping rate of 5 individuals
  - administer genetic test to screen for early-onset Alzheimer’s disease to many individuals
  - identify 5 individuals who might be at risk
  - also administer tapping test to these individuals
- Question: is tapping rate for this group unusually low?
  - N=5; mean = 54.2; standard deviation = 6

Tapping test (group mean)

- Population: μ = 59, σ = 7
- Sample (N=5): mean = 54.2; standard deviation = 6
- Question: is tapping rate for this group unusually low?
  - H₀: μ ≥ 59 (sample was not drawn from Alzheimer’s population)
  - H₁: μ < 59 (sample was drawn Alzheimer’s population)
- Sampling distribution of mean assuming H₀ is true:
  - N(μ,σ²) = N(59, 7²/5) = N(μ=59,σ²=9.8) [σ = sqrt(9.8) = 3.13]
  - z score for our mean: z = (54.2-59) / 3.13 = -1.533
  - significance level = .05 (one-tailed); Critical z = -1.645; p(z ≤ -1.645 | H₀) = 0.05
  - Observed z (-1.533) > Critical z (-1.645)
    - p(z ≤ -1.533 | H₀) = 0.063
    - fail to reject H₀

Testing Hypotheses with Unknown Population Variance
Tapping test (group mean)

- Population: $\mu = 59$, $\sigma = ?$
- Sample (N=5): mean = 54.2; standard deviation = 6
- Question: is tapping rate for this group unusually low?
  - H0: $\mu \geq 59$ (sample was not drawn from Alzheimer’s population)
  - H1: $\mu < 59$ (sample was drawn from Alzheimer’s population)

- Sampling distribution of mean assuming H0 is true:
  - $N(\mu, \sigma^2) = N(59, \sigma^2)$ (assuming $\sigma$ is known)
  - $z$ score for our mean: $z = (54.2 - 59) / 2.68 = -1.79$
  - N.B. This “z” is based on ESTIMATED standard deviation

- Significance level = .05 (one-tailed); Critical $z = -1.645$; $p(z \leq -1.645 | H0) = 0.05$
- Estimated “z” (-1.79) < Critical $z$ (-1.645)
  - $p(z \leq -1.79 | H0) = 0.037$
  - reject H0 in favour of H1

Effect of using estimate of $\sigma$

- $z$ is defined with KNOWN population $\mu$ and $\sigma$
- only source of variation in $z$ is sampling error of mean
- using estimate of $\sigma$ introduces another source of variation in $z$
  - Estimated $z$ depends on group mean and group standard deviation
- does this affect distribution of $z$?

\[ z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}} \]

\[ \hat{z} = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{N}} \]

Effect of using estimate of population variation

- William Gossett applied statistics to his work in the Guinness brewery
- Under the pseudonym, Student, he investigated effects of estimating $\sigma$ on $z$ test
- Discovered that using estimates of $\sigma$ lead to more “extreme” values of $z$ than predicted by statistical theory

Effect of inflating $z$ score

- calculating $z$ with estimated $\sigma$ inflates $z$ scores
- extreme $\hat{z}$ values occur more frequently than expected when H0 is True
- what effect does this have on our evaluation of H0?
Effect of using estimate of population variation

- William Gossett applied statistics to his work in the Guinness brewery
- Under the pseudonym, Student, he investigated effects of estimating $\sigma$ on z test
- Discovered that using estimates of $\sigma$ lead to more “extreme” values of z than predicted by statistical theory
- Caused an increase in Type I errors
  - especially for small samples
- Devised a new test that corrected these errors
  - Student’s t test

$$t \text{ distribution}$$

- unimodal
- symmetrical around zero
- has 1 parameter:
  - degrees of freedom (df)
- df alters kurtosis
  - lower df associated with narrower middle portion & heavier tails
- t approximately normal for df $\geq 35$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

what are “degrees-of-freedom”?

$$\text{degrees of freedom}$$

- consider a set of n=4 numbers:
  - guess the value of each number:
    - 2
    - 0
    - 8
    - 10
- hard to guess correctly because each number can be any value
  - each number is free to vary
degrees of freedom

• consider a set of n=4 numbers, whose total = 20
  - guess the value of each number:
    › 4
    › 1
    › 10
    › 5
  • first 3 numbers can be any value
    - but value of 4th is determined by first 3 (and the total value of 20)
    - 4th value is not free to vary
  • given the total, we say the set of n=4 values has n-1=3 degrees of freedom

— B.S. Everitt, The Cambridge Dictionary of Statistics

degrees of freedom

• consider a set of n=4 numbers, whose mean = 5
  - guess the value of each number:
    › 5
    › 8
    › 6
    › 1
  • first 3 numbers can be any value
    - but value of 4th is determined by first 3 (and the mean value of 5)
    - 4th value is not free to vary
  • given the mean, we say the set of n=4 values has n-1=3 degrees of freedom

• Degrees of freedom show up in many different places in statistics
• when calculating $t$ for 1 sample
  - df = sample size minus one = $n-1$
  - true because $t$ is based on sample variance
  - which depends on sample mean
  - given the mean, only (n-1) sample values are free to vary
    › the nth-value is determined by the other n-1 values

$s^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}$
Back to hypothesis testing

When $\sigma$ is NOT known
- estimated $z$ is inflated
- our standardized score does **not** follow $z$ distribution
- using "$z$" increases Type I error rate

However, standardized score **DOES** follow a $t$ distribution
Therefore, our estimated "$z$" actually is a $t$ statistic
and we use critical values of $t$, not $z$, to evaluate null hypothesis

---

$t = \frac{\bar{X} - \mu}{\hat{\sigma}_X}$

---

- Our tapping test sample: $n=5$, $df=n-1=4$
  - significance level = .05 (one-tailed)
  - **critical $t$ (df=4) = -2.13**
    - $p(t \leq -2.13 \mid H_0) = .05$

---

- Tapping example using $z$ (reminder of key results):
  - significance level = .05 (1-tailed); **Critical $z = -1.645$**; $p(z \leq -1.645 \mid H_0) = 0.05$
  - Observed Estimated $z$ (-1.79) < Critical $z$ (-1.645); reject $H_0$ in favour of $H_1$
    - $p(z \leq -1.79 \mid H_0) = 0.037$; reject $H_0$ in favour of $H_1$
Back to hypothesis testing

- Our tapping test sample: \( n=5, \) \( df=n-1=4 \)
  - significance level = .05 (1-tailed); **critical t**(df=4) = -2.13
  - \( p(t \leq -2.13 | H_0) = .05 \)
- Tapping example using z (reminder of key results):
  - significance level = .05 (1-tailed); **Critical z** = -1.645; \( p(z \leq -1.645 | H_0) = .05 \)
  - Observed Estimated "z" (-1.79) < Critical z (-1.645); reject \( H_0 \) in favour of \( H_1 \)
  - \( p(z \leq -1.79 | H_0) = 0.037 \); reject \( H_0 \) in favour of \( H_1 \)
- Tapping test example using t:
  - **Observed** \( t = (54.2-59) / 2.68 = -1.79; \) \( p(t \leq -1.79 | H_0) = 0.074 \)
  - Observed \( t \) (-1.79) not less than Critical \( t \) (-2.13)
  - so, do **NOT** reject \( H_0 \) in favour of \( H_1 \)

\[
t = \frac{\bar{X} - \mu}{\hat{\sigma}_X}
\]

**two-sided t test (tapping test example)**

- Null & Research Hypotheses:
  - \( H_0: \mu=59 \) (sample drawn from healthy population)
  - \( H_1: \mu \neq 59 \) (sample not drawn from healthy population)
- \( \alpha = .05 \)
  - critical values of \( t = \pm 2.776 \) (df=4, 2-tailed, alpha=.05)
- given \( H_0 \), \( p(t > 2.776) = .025 \) & \( p(t < -2.776) = .025 \)
- sample: \( N=5 \), mean = 54.2, standard deviation = 6
  - following values are the same as for 1-tailed tests:
  - \( t = (54.2-59) / 6(\sqrt{6/5}) = (54.2-59)/2.68 = -1.79 \)
  - **Observed** \( t \) (-1.79) is **not** more extreme than either critical \( t \) value (±2.776)
  - fail to reject \( H_0 \)
**Effect of Sample Size (Tapping Test Example)**

- Null & Research Hypotheses:
  - H0: \( \mu = 59 \) (sample drawn from healthy population)
  - H1: \( \mu \neq 59 \) (sample not drawn from healthy population)
- Sample: \( N = 30 \), mean = 54.2, standard deviation = 6
  - \( t = \frac{54.2 - 59}{\sqrt{6^2/30}} \approx -4.4 \)
- SD of sampling distribution decreases from \( \sqrt{6^2/5} \) to \( \sqrt{6^2/30} \)
  - so same values of \( \mu \), sample mean, and sd result in bigger \( t \) value
- \( \alpha = .05 \) (df=n-1=29)
  - critical values of \( t = \pm 2.04 \) [approximate; value for df=30 taken from Table 12.1]
  - \( p(t < -2.04) = 0.025 \) & \( p(t > 2.04) = 0.025 \)
  - \( p(t < -2.04 \text{ OR } t > 2.04) \mid \text{H0} = 0.025 + 0.025 = 0.05 \)
- Observed \( t \) (−4.0) is more extreme than either critical \( t \) value (±2.04)
  - reject H0 in favour of H1

**Factors Affecting \( t \) Test**

- observed difference between sample mean - \( \mu \) (when H0 is true)
- sample standard deviation (s)
- sample size (N)
  - \( p \)-value depends on N
  - large N means smaller \( p \) values
- significance level (\( \alpha \))
- 1- vs. 2-tailed test

\[
t = \frac{\bar{X} - \mu}{\sigma_X}
\]

**Treatment for Hypertension (High Blood Pressure)**

- measure blood pressure (BP) in 1000 adults
- select 296 adults HBP > 95 mmHg
  - apply treatment
  - measure post-treatment BP
- compare pre- & post-treatment BP
  - \( t \)-test on paired samples
    - each subject gives 2 measures
    - do \( t \) test on difference scores
Treatment for Hypertension (High Blood Pressure)

- Difference: mean = -2.91, sd = 6.02, N = 296
  - H0: True Difference >= 0
  - H1: True Difference < 0
- critical value of t:
  - df = N-1 = 295 & alpha = .05
  - t.critical = -1.65
- observed value of t:
  - t= (-2.91 - 0) / sqrt(6.02^2/296) = -8.31
  - p(t <= -8.31 | H0) < 0.0001
  - reject H0 in favour of H1

- Does that mean our “treatment” worked?
  - Did the treatment cause the reduction in BP?
- Not necessarily!

Is reduction in HBP due to treatment?

What else might cause blood pressure to be lower in the 2nd (post-treatment) test?

- Placebo Effect?
- Increased familiarity, reduced anxiety on 2nd test
- How could we rule out these alternative explanations?
- One more possible explanation:
  - Regression to the Mean

Regression to the Mean

- Very tall parents tend to have shorter children
  - & very short parents tend to have taller children
- Students with very high scores on test 1 tend to have slightly lower scores on test 2
  - & students with very low scores on test 1 tend to have slightly higher scores on test 2
- extreme scores on 1st measurement tend to be closer to average on 2nd measurement

Francis Galton
Regression to the Mean

Note that blood pressure experiment included only subjects with high BP

Is reduction in HBP due to treatment?

What else might cause blood pressure to be lower in the 2nd (post-treatment) test?

- Placebo Effect?
- Increased familiarity, reduced anxiety on 2nd test
- How could we rule out these alternative explanations?
- One more possible explanation:
  - Regression to the Mean
  - Selecting extreme high scores on test 1 means that we should expect a lower average on test 2 even if the drug/treatment has no effect!
  - one way to test this idea: include a no-treatment group

The weird power of the placebo effect, explained Brian Resnick, Vox, 2017
1-sample t tests (summary)

• z tests used when population variance is known
• estimating population variance from sample variance inflates “z”
  ‣ estimated-z is not distributed as standardized normal variable
  ‣ extreme values occur more frequently than expected
  ‣ “z” test has higher Type I error rate than expected
• t tests used when population variance is estimated from sample
  ‣ logic of t test is the same as z test
  ‣ primary difference is we compare observed t to critical value of t, not z
  ‣ corrects for inflation of Type I error rate
  ‣ t test on paired samples is a t test on difference scores
• Be cautious when interpreting results!
  ‣ unusually low or high scores/means, or a difference between pre- & post-treatment scores, may
    occur for several (sometimes non-obvious) reasons!