Concepts from previous lectures

- sampling distributions
  - sampling error
  - standard error of the mean
  - degrees-of-freedom
- Null and alternative/research hypotheses (H0 vs H1)

Tapping test (single observation)

- tapping test from Chapter 8
- tapping rate can be used as an indicator for neurological damage
  - 10 s tapping test in normal pop: mean = 59; sd = 7
  - tapping rate is slower in Alzheimer’s patients (and other clinical populations)
  - measured rate for 1 subject = 45 (per 10 s trial)
- is this case unusual?
General Strategy
reject H0 if z exceeds critical values of z

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 = 5
  - Population: μ = 59, σ = 7
  - Sample (N=5): mean = 54.2; standard deviation = 5
  - 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 = 5
- Question: is tapping rate for this group unusually low?
  - $H_0$: $\mu \geq 59$ (sample was not drawn from Alzheimer’s population)
  - $H_1$: $\mu < 59$ (sample was drawn Alzheimer’s population)
- Sampling distribution of mean assuming $H_0$ is true:
  - $N(\mu, \sigma^2) = N(59, 5/5) = N(\mu=59, \sigma^2=5)$ ($\sigma = \sqrt{5} = 2.23$)
  - $z$ score for our mean: $z = (54.2-59) / 2.23 = -2.152$
    - N.B. This $z$ is based on ESTIMATED standard deviation
    - significance level = .05 (one-tailed); Critical $z = -1.645$; $p(z \leq -1.645 | H_0) = 0.05$
    - “Estimated $z$” (-2.152) < Critical $z$ (-1.645)
      - $p(z < -2.152 | H_0) = 0.0157$
      - reject $H_0$ in favour of $H_1$

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 $z$
  - $z$ score depends on group mean AND group standard deviation
- does this affect distribution of $z$?

Sampling Distribution of Standard Deviation
- standard deviation $s$ has a sampling distribution
- mean of $s = \sigma$
  - $s$ is unbiased estimate of $\sigma$
  - but distribution is skewed
    - median(s) $< \text{mean(s)}$
    - so $s < \sigma$ more than 50% of time
  - what does this mean for $z$?
Under- and over-estimates of $\sigma$

\[
\begin{align*}
z &= \frac{\bar{X} - \mu \bar{X}}{\sigma \bar{X}} \\
\hat{z} &= \frac{\bar{X} - \mu \bar{X}}{\hat{\sigma} \bar{X}}
\end{align*}
\]

1) if $(\hat{\sigma} \bar{X} \geq \sigma \bar{X})$ then $(\hat{z} \leq z)$

2) if $(\hat{\sigma} \bar{X} < \sigma \bar{X})$ then $(\hat{z} > z)$

Skewed sampling distribution of $z$ means that “2” happens more than 50% of the time

Effect of inflating $z$ score

- Calculating $z$ with estimated $\sigma$ inflates $z$ scores
- Extreme $\hat{z}$ values occur more frequently than expected
- What effect does this have on our evaluation of H0?

Distribution of estimated-$z$ is not normal

- Simulation = 10,000 samples
  - $n=5$, $\sigma=7$, $\mu=59$
  - Calculate $z$ for each sample
- Distribution of $z$ has more outliers than expected
- $p(z < -1.645) = 0.092$, not 0.05
  - Type I error rate is higher than expected (.092 vs .05)

William Gosset and Student’s t

- Under the pseudonym, Student, William Gosset investigated effects of estimating $\sigma$ on $z$ test
- Noted that estimated $z$ was not distributed normally
- Identified the correct distribution
  - Student’s $t$ distribution

t 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 $df \geq 35$

what are “degrees-of-freedom”?  

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
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 (df)

- 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

---

Distribution of estimated-z is not normal

- Simulation = 10,000 samples
  - tapping test parameters:
    - n=5, \( \sigma=7, \mu=59 \)
  - calculate estimated-z for each sample using sample standard deviation
  - Distribution of estimated-z has more outliers than expected
- \( p(“z” < -1.645) = 0.092, \) not 0.05
  - Type I error rate is higher than expected (.092 vs .05)
Back to hypothesis testing

- when \( \sigma \) is NOT known
  - estimated \( z \) is inflated
  - standardized score is distributed as \( t \), not \( z \)
  - using \( z \) causes Type I error rate to be larger than the expected value

- solution: use critical values of \( t \), not \( z \)

Our tapping test sample: \( n=5 \), \( df=n-1=4 \)
- significance level = .05 (one-tailed); critical \( t(df=4) = -2.13 \)
- \( p(t \leq -2.13 \mid H_0) = .05 \)
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 \( z = -1.645 \); \( p(z \leq -1.645 | H_0) = 0.05 \)
  - Observed Estimated-\( z \) (-2.152) < Critical \( z \) (-1.645); reject \( H_0 \) in favour of \( H_1 \)
  - \( p(z \leq -2.152 | H_0) = 0.0157 \); reject \( H_0 \) in favour of \( H_1 \)

- Tapping test example using \( t \):
  - Observed \( t \) = \([54.2 - 59] / 2.23 = -2.152; \( p(t \leq -2.152 | H_0) = 0.049 \)
  - Observed \( t \) (-2.152) < Critical \( t \) (-2.13) (close!)
  - still reject \( H_0 \) in favour of \( H_1 \)

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

- Null & Research Hypotheses:
  - \( H_0: \mu=59 \) (sample drawn from healthy population)
  - \( H_1: \mu\neq59 \) (sample not drawn from healthy population)
- \( \alpha = .05 \)
  - critical values of \( t = \pm 2.776 \) [df=4, 2-tailed, \( \alpha=.05 \)]
two-sided t test (tapping test example)

- Null & Research Hypotheses:
  - H0: μ=59 (sample drawn from healthy population)
  - H1: μ≠59 (sample not drawn from healthy population)
- alpha = .05
  - critical values of t = ± 2.776 (df=4, 2-tailed, alpha=.05)
  - given H0, p(t > 2.776) = .025 & p(t < -2.776) = .025
- sample: N=5, mean = 54.2, standard deviation = 5
  - following values are the same as for 1-tailed tests:
    - t = (54.2-59) / sqrt(5/5) = (54.2-59)/2.23 = –2.152
  - Observed t (-2.152) is not more extreme than either critical t value (±2.776)
    - fail to reject H0

effect of sample size (tapping test example)

- Null & Research Hypotheses:
  - H0: μ=59 (sample drawn from healthy population)
  - H1: μ≠59 (sample not drawn from healthy population)
- sample: N=30, mean = 54.2, standard deviation = 5
  - t = (54.2-59) / sqrt(5/30) = (54.2-59)/0.913 = –5.257
  - SD of sampling distribution decreases from sqrt(5/5) to sqrt(5/30)
  - so same values of μ & sample mean result in bigger t value
- alpha = .05 (df=n-1=29)
  - critical values of t = ± 2.04 [approximate; value for df=30 taken from Table 12.1]
  - given H0, p(t > 2.04) = .025 & p(t < -2.04) = .025
  - Observed t (–5.257) is more extreme than either critical t value (±2.04)
    - reject H0 in favour of H1

Effect of Sample Size on SEM

\[ \hat{\sigma}_X = \frac{s}{\sqrt{N}} \]

Factors affecting t test

- observed difference: (sample mean - μ)
- sample standard deviation (s)
- sample size (N)
  - p-value depends on N
  - reason for calculating effect size
    - less dependent on N
- significance level (α)
- one- vs. two-tailed test
1-sample t tests (summary)

• z tests are 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
  - causes our z test to have a Type I error rate that is higher than the expected value (i.e., alpha)
• t tests correct for this inflation of Type I error rate
  - “estimated-z” follows Student’s t distribution
  - logic of t test is the same as z test
  - primary difference is we compare observed “t” to critical value of “t”, not z