PSYC 5301 - Lecture 1

Today's goal - review of classical (frequentist) statistical inference

Example: Suppose we are testing a treatment that has been proposed to increase intelligence (as measured by IQ)

A sample of $N = 25$ people is given the treatment, and the average IQ for the sample is $\bar{X} = 107$.

Did the treatment work?

We can answer by translating this research question to a statistical question.

Let μ = mean of the population who receive treatment

Is $\mu > 100$, the average IQ for the general population?

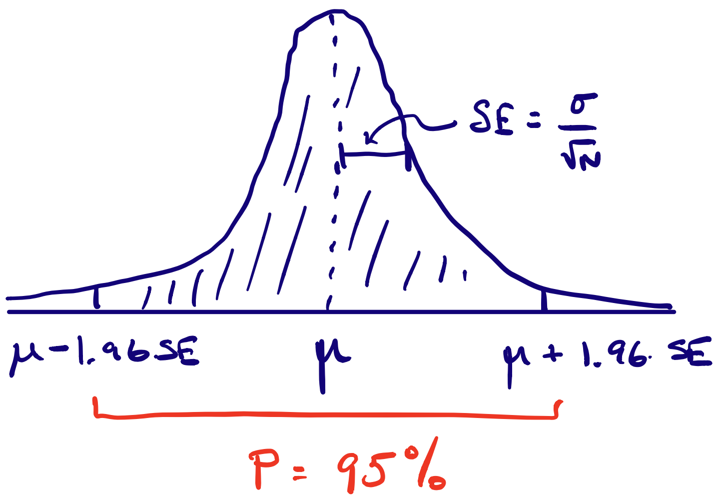
Two methods:

(1) estimate μ from $\bar{x}$ (parameter estimation)

(2) test competing hypotheses about μ.
(model comparison)

Parameter Estimation – what is the value of μ?

From earlier courses, we know that 95% of sample means are within (almost) two standard deviations of the population mean μ.



So, there is a 95% probability that any given sample mean is between $\mu - 1.96 \cdot SE$ and $\mu + 1.96 \cdot SE$

$$\hookrightarrow \quad \mu - 1.96 \cdot \frac{\sigma}{\sqrt{N}} \leq \bar{x} \leq \mu + 1.96 \cdot \frac{\sigma}{\sqrt{N}}$$

A little algebra converts this to:

$$\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{N}} \leq \mu \leq \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{N}}$$

<u>Definition:</u> A <u>95% confidence interval</u> for μ is given by the interval

$$\left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{N}}, \ \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{N}}\right)$$

Back to our example: recall that our treatment sample ($N = 25$) had a mean of $\bar{X} = 107$. Let's compute a 95% confidence interval (CI) for μ.

Recall: for distribution of IQ scores, we know $\sigma = 15$.

So

$$\begin{aligned}
95\%\,CI &= \left(\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{N}}, \ \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{N}}\right) \\
&= \left(107 - 1.96 \cdot \frac{15}{\sqrt{25}}, \ 107 + 1.96 \cdot \frac{15}{\sqrt{25}}\right) \\
&= \left(107 - 5.88, \ 107 + 5.88\right) \quad \leftarrow \text{Margin of error} \\
&= (101.12, \ 112.88).
\end{aligned}$$

So, we are 95% confident that μ is between 101.12 and 112.88

<u>Note:</u> Since our estimate for μ is greater than 100, we can conclude that the treatment worked.

Hypothesis testing

We define two competing hypotheses (models) about μ

$H_0: \mu = 100$ ("null hypothesis" / no tmt effect)

$H_1: \mu > 100$ ("alternative hypothesis" / positive tmt effect)

Let us assume that H_0 is true (that is, $\mu = 100$)

What is the probability of observing our sample mean $\bar{x} = 107$ (or more extreme) if H_0 is true?

$p(\bar{x} \geq 107)$ ← translate to z-score

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{N}} = \frac{107 - 100}{15/\sqrt{25}} = \frac{7}{15/5} = \frac{7}{3} = 2.33.$$

i.e.

$P(z \geq 2.33) = 0.0099$ ← "p-value"

Conclusion: our data is rare under H_0.

So, we reject H_0 as a plausible hypothesis.

This gives support for $H_1: \mu > 100$.

Technical note – you may have noticed in our example that we knew the population standard deviation σ.

What happens if we are not given σ?

<u>Example:</u> A population has a mean of 23. A sample of $N=4$ is given an experimental treatment and had scores of 20, 22, 22, and 20. Does the treatment result in a significantly lower score?

We are not given σ – what can we do?

How do we get σ?

- maybe we can estimate it from the observed data

$$\hookrightarrow \quad s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$= \sqrt{\frac{SS}{N}}$$

- Problem: s tends to be too small! It systematically underestimates σ.

• Solution: let's correct the formula to fix the bias

"sigma hat"

$$\hat{\sigma} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{SS}{N-1}}$$

"sample standard deviation"

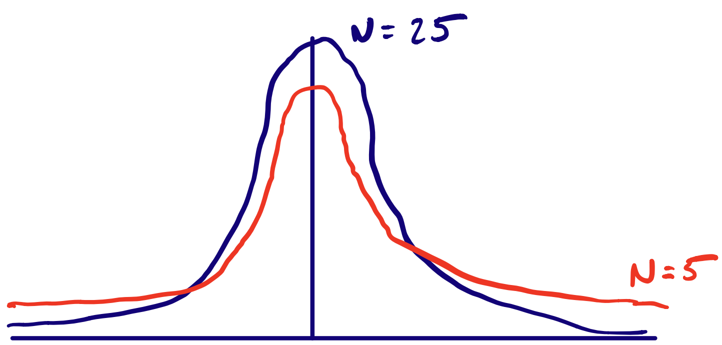
OK, fine ... but the distribution of "z-scores" $\left(\frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{N}} \right)$ is no longer normal, but something else entirely!

The T - distribution

- details worked out by Gosset (1908)
  - ↳ Biometrika paper written under pseudonym "Student"
  - ↳ nice history of this paper given in Zabell (2008) - Journal of the Amer. Stat. Assoc.

- shape of the distribution depends on sample size
  - ↳ parameter = "degrees of freedom"
  - ↳ $df = N - 1$

- the smaller the sample size, the fatter the tails



- so, sample size (i.e., degrees of freedom) must be specified when we calculate probabilities

Back to our example:

A population has a mean of 23. A sample of $N=4$ is given an experimental treatment and had scores of 20, 22, 22, and 20. Does the treatment result in a significantly lower score?

Let μ = mean of the treatment population. Note that $\bar{X} = 21$.

Define: $H_0: \mu = 23$

$H_1: \mu < 23$.

Assume H_0 is true.

Find probability of observing $\bar{X} < 21$ if H_0 is true.

To proceed, we need to compute an estimate $\hat{\sigma}$ of the population standard deviation.

- from above, we have $\hat{\sigma} = \sqrt{\frac{SS}{N-1}}$

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
20	-1	1
22	1	1
22	1	1
20	-1	1

$\bar{X} = 21$

$\longrightarrow SS = 4$

$\longrightarrow \hat{\sigma} = \sqrt{\frac{SS}{N-1}}$

$= \sqrt{\frac{4}{3}} = 1.15$

Now we can compute a "t - score"

$$t = \frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{N}} = \frac{21 - 23}{1.15/\sqrt{4}} = \frac{-2}{0.575} = -3.48$$

Finally, we find $P(t < -3.48)$

↳ from app (with $df = 4 - 1 = 3$)

we get $P = 0.02$

Since $P < 0.05$, our data is rare if H_0 is true.

So we reject H_0 in favor of H_1 (i.e, $\mu < 23$)

and conclude that the treatment results in significantly lower scores.

What about confidence intervals?

Recall: we can estimate a 95% confidence interval for an unknown population mean μ by using the sample mean $\bar{X}$ and the population standard deviation σ as

$$\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{N}}$$

What if we are not given σ?

Can we use our estimate $\hat{\sigma} = \sqrt{\frac{SS}{N-1}}$?

Well, yes – sort of – but we have to adjust the 1.96

Why?

* 1.96 is used because for a normal distribution, 95% of sample means fall between $-1.96 \cdot SE$ and $1.96 \cdot SE$. This assumes σ is known.

* if estimating σ with $\hat{\sigma}$, we get a t-distribution for the sample means. The exact shape of this distribution depends on the size of the sample.

In light of this, let's define a generalized confidence interval

$$\overline{X} \pm t^*_{df} \cdot \frac{\hat{\sigma}}{\sqrt{N}}$$

where the value of t^*_{df} depends on sample size

→ defined as the value of t which leaves 5% of the distribution in the two tails (combined).

↳ sometimes called the critical value of the t-distribution

↳ easy to find from distribution calculator app

Back to our example: earlier we found $\overline{X} = 21$

$\hat{\sigma} = 1.15$

From app, we find $t^*_{df} = 3.18$

So 95% CI $= \overline{X} \pm t^*_{df} \cdot \frac{\hat{\sigma}}{\sqrt{N}}$

$= 21 \pm 3.18 \cdot \frac{1.15}{\sqrt{4}}$

$= 21 \pm 1.83$

$= (19.17, 22.83)$

Take home:

* translate research questions to statistical questions about some population parameter (e.g., μ)

* Estimation – compute 95% confidence interval for μ

* Hypothesis testing – define competing hypotheses about μ (H_0, H_1)

    - assume H_0 is true
    - if observed data is rare under H_0, we reject H_0 and conclude support for H_1.

* in problems where σ is unknown, we must estimate it from the data.

    - when using estimate $\hat{\sigma}$, the distribution of sample means <u>depends on sample size</u>.

    - result: t-test.