Sets 25-27 Part A: Hypothesis Testing

Stat 260: July 19, 2024

Example 1: *A city sets its maximum acceptable nickel content in soil for residential regions at 45.0 mg/kg. A series of 80 soil samples are taken from around the region, and an average nickel content of 45.6 mg/kg and standard deviation of 3.1 mg/kg are found. Is there evidence to suggest the mean soil nickel content exceeds the city's acceptable limit (45.0 mg/kg)?*

Law Example: *In the Canadian justice system, a suspect is presumed innocent until proven guilty.*

- To answer the above question, we try to decide between two hypotheses:
  (a) ***Null Hypothesis*** H_0***:*** The hypothesis of "no change"; what is assumed to be true.

  Ex1: mean Nickel Soil content is within safe limit

  Law: the suspect's innocent

  (b) ***Alternative Hypothesis*** H_1***:*** The hypothesis of change.

  Ex1: The mean nickel soil content exceeds safe limit

  Law: the suspect is guilty.

- In the *p****-value approach to hypothesis testing***, we calculate the *p****-value***, which the probably of seeing a result as extreme (i.e. as big or as small) as the observed sample, *assuming that the null hypothesis is true.*

  Ex 1: p-value = probability of finding 80 soil samples with a mean nickel content as high as 45.6 mg/kg if H_0 were true

  - If the p-value is large, our observations are consistent with the null hypothesis.

    ↳ It's likely we could have found a sample that extreme

    So our assumption H_0 is true (its safe) is probably reasonable

  - If the p-value is small, then our observations are **not** consistent with the null hypothesis.

    ↳ Its likely that our original assumptions were incorrect. (H_0 is true)

- Two outcomes of hypothesis testing:
  (i) There is enough evidence to reject H_0. ↗ small p-value

  "reject H_0" and take H_1 in its place

  (ii) Not enough evidence to reject H_0.

  "fail to reject H_0"

  Never: prove H_0 is true

  we are gathering evidence against H_0 not for it

The p-value Approach to Hypothesis Testing:

1. Define the parameter(s) to be tested.
2. Define the null (H_0) and the alternative (H_1) hypotheses.
3. Specify the test statistic and the distribution under H_0.
4. Find the observed value of the test statistic.
5. Calculate the p-value.
6. Report the strength of evidence against H_0.
   - ▷ *"Very strong evidence"* if p-value ≤ 0.01
   - ▷ *"Strong evidence"* if $0.01 < p$-value ≤ 0.05
   - ▷ *"Moderate evidence"* if $0.05 < p$-value ≤ 0.10
   - ▷ *"Little to no evidence"* if $0.10 < p$-value.
7. Answer any other questions given (ex. report the value of estimate and/or the estimated standard error, etc).

Example 1 Continued...

1. Define the parameter(s) to be tested.

μ = population mean nickel soil content

2. Specify H_0 and H_1. (always in terms of parameter)

$\underbrace{H_0 = \mu \leq 45 \text{ mg/kg}}_{\text{soil is within safe limits}}$ $\qquad$ $\underbrace{H_1 = \mu > 45 \text{ mg/kg}}_{\text{soil exceeds safe limits}}$

3. Specify the test-statistic to be used, and identify its distribution (assuming H_0 is true).

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim Z \qquad \frac{\bar{X} - \mu}{S/\sqrt{n}} \begin{array}{l} \sim Z \text{ (if } n \geq 30) \\ \sim T_r \text{ (if } n < 30) \end{array}$$

same as confidence interval choices

4. Compute the observed value of the test-statistics.

Since $n = 80 > 30$ and σ is unknown use:

value from H_0

$$Z_{\text{observed}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{45.6 - 45}{3.1/\sqrt{80}} = 1.73$$

5. Compute the p-value. p-value $= P(\bar{x} > 45.6 \mid H_0 \text{ is true})$ ↗ dont need to show in an exam

$$\underbrace{P\left(\frac{\bar{x}-\mu}{s/\sqrt{n}} > \frac{45.6-45}{3.1/\sqrt{80}}\right)}_{\text{dont need to show in an exam}} = P(Z > Z_{observed}) = P(Z > 1.73)$$

$$= 1 - P(Z < 1.73) = 1.09582 = 0.0418$$

6. Report the strength of evidence against H_0.

Strong evidence against H_0.

Strong evidence to suggest mean nickel soil level is not withis safe limits.

7. (*If asked*) Report the estimated value of the parameter and (estimated) standard error (ese or se).

estimated value of parameter: $\bar{x} = 45.6$ mg/kg

estimated standard error: $\frac{s}{\sqrt{n}} = \frac{3.1}{\sqrt{80}} = 0.347$ mg/kg

- **Type I Error:** Evidence leads us to reject H_0, but H_0 is actually true.
  - ▷ Occurs with probability α (the ***significance level***).

    Ex1: Conclude soil is unsafe but it actually is safe

    Law: we convict an innocent

- **Type II Error:** Evidence leads us to not reject H_0, but H_0 is actually false.
  - ▷ Occurs with probability β (where $1 - \beta$ is the ***power***).

    Ex1: conclude soil is safe when it actually exceeds safe limits

    Law: Let guilty person go free.

- Often a decrease in α results in an increase in β - but increasing sample size can reduce the chance of both types of errors.

Significance Level Approach to Hypothesis Testing

- Before sampling and testing, α is specified by the user/investigator (typically $\alpha = 0.1$, 0.05, or 0.01), to give control over wrongly rejecting H_0.
- (*If asked*) Test H_0 at the provided significance level α.

  If p-value $\leq \alpha$, reject H_0.

  If p-value $> \alpha$, do not reject H_0.

Example 1 Continued...

- Test the hypothesis at $\alpha = 0.05$.

p-value $= 0.0418 < 0.05 = \alpha$
since p-value $< \alpha$
we reject H_0

There is sufficient evidence to suggest that the nickel soil content exceeds safe levels (45 mg/kg)

- Test the hypothesis at $\alpha = 0.01$.

p-value $= 0.0418 > 0.01 = \alpha$
since p-value $> \alpha$
we fail to reject H_0

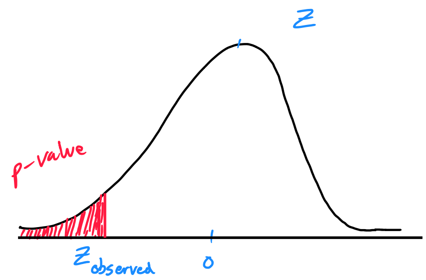
There isn't sufficient evidence to suggest that the nickel soil content exceeds safe levels (45 mg/kg)
It's reasonable to assume that the nickel content is within safe levels.

WARNING: It is statistically dishonest to set the value of α after the data has been collected. The value of α should be set taking into account the consequences of Type I and Type II errors *before* the study is carried out.

Three Types of Hypothesis Tests

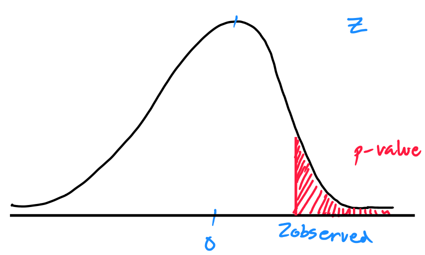
μ_0 = value of μ from H_0

1. **Left-tailed Test**
   - $H_0 : \mu = \mu_0$ or $H_0 : \mu \geq \mu_0$.
   - $H_1 : \mu < \mu_0$
   - p-value $= P(Z < z_{obs})$



Literally left tail

2. **Right-tailed Test**
   - $H_0 : \mu = \mu_0$ or $H_0 : \mu \leq \mu_0$.
   - $H_1 : \mu > \mu_0$
   - p-value $= P(Z > z_{obs}) = 1 - P(Z < z_{obs})$

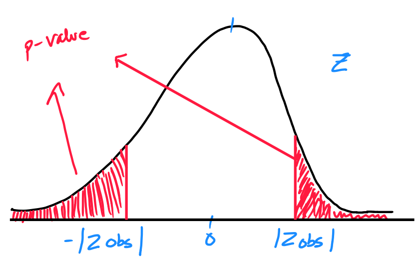


Literally right tail

3. **Two-tailed Test**
   - $H_0 : \mu = \mu_0$
   - $H_1 : \mu \neq \mu_0$
   - p-value $= 2P(Z < -|z_{obs}|)$

p-value = 2(left tail)

* Z_{obs} maybe positive or negative, depending on the data



Example 1 Continued...

Repeat Example 1 with the research question: "Is there evidence to suggest the mean soil nickel content is not 45mg/kg?" $H_1: \mu \neq 45$

Recall that a series of 80 soil samples are taken from around the region, and an average nickel content of 45.6 mg/kg and standard deviation of 3.1 mg/kg are found.

1. Define the parameter(s) to be tested.

same

μ = population mean nickel soil content (mg/kg)

2. Specify H_0 and H_1.

$H_0 = \mu = 45$ mg/kg

$H_1 = \mu \neq 45$ mg/kg

3. Specify the test-statistic to be used, and identify its distribution (assuming H_0 is true).

same

σ unknown, $n = 80 \geq 30$

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

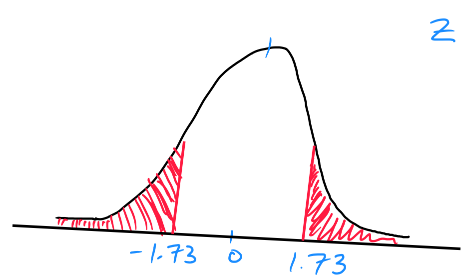
same

4. Compute the observed value of the test-statistics.

$$Z_{obs} = \frac{45.6 - 45}{3.1/\sqrt{80}} = 1.73$$

5. Compute the p-value.

$$p\text{-value} = 2P(Z < -|Z_{obs}|)$$
$$= 2 \text{ (left tail)}$$
$$= 2P(Z < -1.73)$$
$$= 2\underbrace{(0.0418)}_{\text{from table}} = 0.0836$$



6. Report the strength of evidence against H_0.

Moderate evidence against H_0.

Moderate evidence to suggest nickel soil content is not 45 mg/kg

Extra Example 1: In salmon-bearing streams, wildlife like bears and eagles distribute the nutrient-rich remains of fish to the forest floor. Suppose that a study examines 50 trees within 10m of salmon-bearing streams and finds a sample mean growth rate of 15.4cm per year, with a sample standard deviation of 1.3 cm per year.

(a) Is there evidence to suggest that trees growing within 10m of a salmon-bearing stream have an average growth rate of greater than 15cm per year?

1. Define the parameter(s) to be tested.

2. Specify H_0 and H_1.

3. Specify the test-statistic to be used, and identify its distribution (assuming H_0 is true).

4. Compute the observed value of the test-statistics.

5. Compute the p-value.

6. Report the strength of evidence against H_0.

(b) Repeat the previous example, but with the research question: "*Is there evidence to suggest that trees growing within 10m of a salmon-bearing stream do not have an average growth rate of 15cm per year?*"

Example 2: The giant Pacific octopus (*Enteroctopus dofleini*) is noted for being long-lived compared to many other octopus species. Suppose that a certain wild population is known to have normally distributed lifespans, with an average of 4 years. A researcher wishes to test to the claim that captive octopuses have a different lifespan. The lifespans (in years) of 6 captive giant Pacific octopuses are recorded below.

$\bar{x} = 4.78$

$s = 0.96$

3.4	4.8	5.6	3.9	5.8	5.2

(a) Test the claim that the average lifespan of captive giant Pacific octopus $\underbrace{\text{is not 4 years.}}_{\hookrightarrow \text{2 tail test}}$

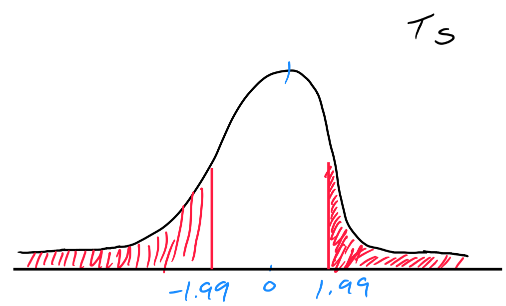
Let μ = population mean lifespan of captive giant pacific octopuses

$H_0 = \mu = 4$ years $\quad H_1 = \mu \neq 4$ years

σ is unknown and $n = 6 < 30$ so T_γ test statistic:

$T = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} \sim T_\gamma$ where $\gamma = n-1 = 6-1 = 5$

$\gamma = 5$



observed value of test statistic:

$t_{obs} = \dfrac{4.78 - 4 \nearrow H_0}{0.96/\sqrt{6}} = 1.99$

T_γ dist gives only right tails so:

p- value:

p-value = 2(right tail)

$= 2P(T_5 > 1.99)$

$(2)\, 0.05 < (2)\, P(T_5 > 1.99) < (2)\, 0.1$

$0.10 <$ p-value < 0.2

Conclusion:

Little to no evidence against H_0

No evidence to suggest the mean lifespan is not 4 years.

(b) Test the claim that the average lifespan of captive giant Pacific octopus is greater than 4 years.

[*Ans:* $0.05 < p - value < 0.10$*: Moderate evidence against* H_0]

(c) Create the 90% CI for the population mean lifespan.

[*Ans:* $[3.99, 557]$]

$\alpha < $ p-value

Relationship to Confidence Interval Estimation:

Testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq$ [redacted]$_0$ at significance level α is equivalent to testing computing a $(1 - \alpha)100\%$ confidence interval for μ, where:

- ▷ we reject H_0 if μ_0 is outside the confidence interval.
- ▷ we fail to reject H_0 if μ_0 is in the confidence interval.

Example 2 Continued... From part c : 90% CI is [3.99, 5.57] years

Hypothesis test for μ:

$H_0 : \mu = 4$ $\quad H_1 : \mu \neq 4$

Since 4 is in the interval [3.99, 5.57], this is equivalent to showing p-value > $\alpha = 0.1$

Conclusion:

At significance level $\alpha = 0.1$, we fail to reject H_0

Summary of single sample inference based on random samples [1] Which test to use can be decided using below table

Sample Data	Test Statistic	Comments
Normal, σ known.	$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	σ rarely known.
Normal, σ unknown, $n < 30$.	$\dfrac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$	
Dist unspecified, σ known, $n \geq 30$	$\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	σ rarely known. Based on CLT.
Dist unspecified, σ unknown, $n \geq 30$	$\dfrac{\overline{X} - \mu}{s/\sqrt{n}} \sim N(0,1)$	Based on CLT
Binomial	Binomial	
Binomial, $np \geq 5$, $n(1-p) \geq 5$	$\dfrac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$	Based on Normal approx. to Binom

Readings and Practice problems: See PDF on Brightspace for detailed instructions.

[1] From Swartz *Introduction of Probability and Statistics* pg 133