Chapter 5

Recap Sheet

• What is a hypothesis test ?

When two competing claims about a parameter are plausible, we use a hypothesis test to decide which claim the data support.

Parts :

1. Parameter of interest (ex : the mean burning rate μ (cm / s) of a propellant)

2. Null hypothesis H_0 : the "status - quo" claim we test against. We always write it with equality (possibly $\leq$ or $\geq$) (ex : $H_0 : \mu = 50$ cm / s)

3. Alternative hypothesis H_1 : the competing claim we hope to find evidence for. It can be :
   - Two - sided (ex . $H_1 : \mu \neq 50$ (mean is different, either higher or lower))
   - One - sided (lower - tailed) (ex : $H_1 : \mu > 50$ (concern is "too low"))
   - One - sided (upper - tailed) (ex : $H_1 : \mu < 50$ (concern is "too high"))

Convention : if H_1 uses $>$ → H_0 uses $\leq$

if H_1 uses $<$ → H_0 uses $\geq$

if H_1 uses $=$ → H_0 uses $=$

What we do with data :

- Take a random sample
- Compute a test statistic (ex : a standardized sample mean)
- Decide wheter the statistic is "too extreme" for H_0
  - → if yes , reject H_0 in favor of H_1
  - → otherwise, fail to reject H_0 (we do not prove H_0 is true)

• Test statistic , significance level and critical region

Test statistic : A test statistic is a single number you compute from the sample that summarizes the evidence against H_0.

- For a mean with known σ, we use the standardized mean :

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

- For a mean with unknown σ, we'll later use a t -statistic.

Under H_0 (and the model assumptions), we know the reference distribution of the test statistic (ex : $Z_0 \sim \mathcal{N}(0,1)$). That lets us judge whether the observed value is "too extreme".

Significance level α and the decision rule.

- You choose a small α (often 0.05)
- α is the probability of a Type I error (rejecting a true H_0)
- Using α, we set up a critical region (values that lead to rejecting H_0) and an acceptance region (fail to reject).

Two-sided test for a mean (σ known):

- Hypotheses : $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
- Critical values for the z-test are $\pm z_{1-\frac{\alpha}{2}}$.
  - z-form rule : Reject H_0 if $|Z_0| \geq z_{1-\frac{\alpha}{2}}$
  - $\bar{x}$-form rule : Reject if $\bar{X} \leq \mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ or $\bar{X} \geq \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

  Otherwise, fail to reject H_0.
  The interval $[\mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$ is the acceptance region for $\bar{X}$.

One-sided tests:

- Upper-tailed $H_1 : \mu > \mu_0$: reject if $Z_0 \geq z_{1-\alpha}$
  (or $\bar{X} \geq \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$)
- Lower-tailed $H_1 : \mu < \mu_0$: reject if $Z_0 \leq z_\alpha$
  (or $\bar{X} \leq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$, noting $z_\alpha < 0$)

Ex : Given $\mu_0 = 50$, $\sigma = 2$, $n = 25$, $\alpha = 0.05$ (two-sided).

- $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$
- Acceptance region for $\bar{X}$: $\mu_0 \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 50 \pm 1.96 \times \frac{2}{\sqrt{25}}$
  $= 50 \pm 0.784$

So accept H_0 if $49.216 \leq \bar{X} \leq 50.784$, otherwise reject.

• Type I and Type II errors, significance α and power

Four possible outcomes : When we test H_0 vs H_1 :

Decision	H_0 true	H_0 false
Fail to reject H_0	Correct	Type II error (β)
Reject H_0	Type I error (α)	Correct

- **Type I error** : reject a true H_0
  $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$
  You choose α (ex : 0.05). It sets the critical region.

- **Type II error** : fail to reject when H_0 is false
  $\beta(\theta) = P(\text{fail to reject } H_0 \mid \text{true parameter} = \theta \in H_1)$
  β depends on the true effect size, n, and variability.

- **Power** : probability to detect a false H_0.
  $\text{Power}(\theta) = 1 - \beta(\theta) = P(\text{reject } H_0 \mid \theta \in H_1)$

How to compute β / power :
1. Determine the acceptance region under your chosen α
2. Under a specific alternative (ex : $\mu = \mu_1$), compute the probability that the test statistic (or $\bar{x}$) lands inside that acceptance region - that's $\beta(\mu_1)$
3. Power at μ_1 is $1 - \beta(\mu_1)$

Ex : Acceptance region $[49.216, 50.784]$

If the true mean were $\mu = 52$, then $\bar{X} \sim \mathcal{N}(52, \frac{\sigma^2}{n}) = \mathcal{N}(52, 0.16)$

$$\beta(52) = P(49.216 \leq \bar{x} \leq 50.784) = \Phi\left(\frac{50.784 - 52}{0.4}\right) - \Phi\left(\frac{49.216 - 52}{0.4}\right)$$
$$\approx \Phi(-3.04) - \Phi(-6.96)$$
$$\approx 0.0012$$

So power at $\mu = 52$ is ≈ 0.9988 (very high because the shift is large relative to the SE)

- **P-values and how they relate to fixed-α tests**

What is a P-value : Given your observed test statistic, the P-value is the probability - assuming H_0 is true - of getting a result at least as extreme (in the direction of H_1) as what you saw.

It answers : "How surprising is my data if H_0 were true ?"
- Decision rule : Reject H_0 if P-value $\leq \alpha$
- Equivalent view : The P-value is the smallest significance level at which you would reject H_0.

Formulas for the z-test on a mean (σ known) :

Let $z_0 = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ and let $\Phi(\cdot)$ be the standard normal CDF.

- Two-sided H_1 : $\mu \neq \mu_0 \longrightarrow p = 2[1 - \Phi(|z_0|)]$
- Upper-tailed H_1 : $\mu > \mu_0 \longrightarrow p = 1 - \Phi(z_0)$
- Lower-tailed H_1 : $\mu < \mu_0 \longrightarrow p = \Phi(z_0)$

Ex : Test $H_0 : \mu = 50$ vs $H_1 : \mu \neq 50$ with $\sigma = 2, n = 25, \bar{x} = 51.3$
- Test statistic : $z_0 = \dfrac{51.3 - 50}{2/\sqrt{25}} = \dfrac{1.3}{0.4} = 3.25$
- Two-sided P-value : $p = 2[1 - \Phi(3.25)] \approx 2(1 - 0.9994)$
  ≈ 0.0012

- Decision : $p = 0.0012 \leq 0.05 \to$ reject H_0 (also reject at 1%)

Interpretation : data this extreme (or more) would occur only about 0.12 % of the time if $\mu = 50$.

- ## General procedure for a hypothesis test

1. Parameter of interest : State clearly what you are testing (ex : μ, σ^2, p) and the measurement units / context

2. Null hypothesis H_0 : The status-quo / equality claim (ex : $H_0 : \mu = \mu_0$). It's the model you assume when computing probabilities.

3. Alternative hypothesis H_1 : What you want evidence for. Choose two-sided ($\neq$) or one-sided ($>$ or $<$) to match the practical question.

4. Test statistic (and model assumptions) . Pick the statistic whose reference distribution under H_0 you know.
   - Mean, σ known $\to Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
   - Mean, σ unknown & Normal $\to T_0 = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ with $df = n-1$
   - Variance (Normal) $\to \dfrac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$
   - Proportion (large n) $\to Z_0 = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

5. Reject / fail rule : Choose your significance level α and either
   - set critical value / region, or
   - plan to report a P-value and compare to α

6. Compute : Plug in the data to get the test-statistic and the P-value (or compare to the critical values)

7. Conclude in context :
   - Statistical decision : "Reject H_0" or "Fail to reject H_0" at level α
   - Practical interpretation

- ## One-sample z-test for a mean when σ is known

- Assumptions : random sample $X_1, \dots, X_n$, population is Normal or n is large (CLT)

- Hypotheses :
  - Two-sided : $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
  - Upper-tailed : $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$
  - Lower-tailed : $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$

- Test statistic (under H_0) : $Z_0 = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \overset{H_0}{\sim} \mathcal{N}(0,1)$
- Decision rules at significance level α :
  - Using critical values :
    - Two - sided : reject H_0 if $|Z_0| \geq z_{1-\frac{\alpha}{2}}$
    - Upper - tailed : reject if $Z_0 \geq z_{1-\alpha}$
    - Lower - tailed : reject if $Z_0 \leq z_\alpha$ (remember $z_\alpha < 0$)
  - Using the P- value :
    - Two - sided : $p = 2[1 - \Phi(|Z_0|)]$
    - Upper - tailed : $p = 1 - \Phi(Z_0)$
    - Lower - tailed : $p = \Phi(Z_0)$

    Reject H_0 if $p \leq \alpha$

Ex : A propellant has specification $\mu_0 = 50$ cm/s. From $n = 25$ tests you get $\bar{x} = 49.6$ cm/s. Historical $\sigma = 2$ cm/s is known. Test : $H_0 : \mu = 50$ vs $H_1 : \mu \neq 50$ at $\alpha = 0.05$.

1. Compute the test statistic : $Z_0 = \dfrac{49.6 - 50}{2/\sqrt{25}} = -1.00$

2. Critical - value decision : $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$.

Since $|Z_0| = 1.00 < 1.96$, fail to reject H_0.

3. P - value (two - sided) : $p = 2[1 - \Phi(|-1.00|)$
$= 2[1 - 0.8413]$
$= 0.3174$

$p = 0.3174 > 0.05 \rightarrow$ fail to reject H_0

4. Conclusion : At the 5 % level, the data do not provide evidence that the mean burning rate differs from 50 cm/s.

- One sample t - test for a mean when σ is unknown

- Assumptions : random sample $X_1, \dots, X_n$, population is Normal or n is large (CLT)
- Hypotheses :
  - Two – sided : $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
  - Upper - tailed : $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$
  - Lower - tailed : $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$
- Test statistic (under H_0) : $T_0 = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}} \overset{H_0}{\sim} t_\nu$, $\nu = n - 1$

Here S is the ssd and t_ν is the Student t distribution with ν degrees of freedom

- Decision rules at significance level α :
  - Using critical values :
    - Two - sided : reject H_0 if $|T_0| \geq t_{1-\frac{\alpha}{2}, \nu}$
    - Upper - tailed : reject if $T_0 \geq t_{1-\alpha, \nu}$
    - Lower - tailed : reject if $T_0 \leq t_{\alpha, \nu}$ (note $t_{\alpha,\nu} < 0$)
  - Using the P- value :
    - Two - sided : $p = 2[1 - F_{t,\nu}(|T_0|)]$
    - Upper - tailed : $p = 1 - F_{t,\nu}(T_0)$
    - Lower - tailed : $p = F_{t,\nu}(T_0)$

    Reject H_0 if $p \leq \alpha$

Ex : A process has target $\mu_0 = 50$. You take $n = 25$ observations and find $\bar{x} = 49.6$, sample SD $s = 2.2$.

Test $H_0 : \mu = 50$ vs $H_1 : \mu \neq 50$ at $\alpha = 0.05$.

1. Test statistic : degrees of freedom $\nu = 24$

$$T_0 = \frac{49.6 - 50}{2.2 / \sqrt{25}} = -0.909$$

2. Critical - value decision : $t_{1-\frac{\alpha}{2}, \nu} = t_{0.975, 24} = 2.064$

Since $|T_0| = 0.909 < 2.064 \rightarrow$ fail to reject H_0.

3. P- value : with $\nu = 24$, $p = 2[1 - F_{t,24}(0.909)] = 0.37 > 0.05$
$\rightarrow$ fail to reject H_0

Conclusion : At 5% level, the data do not provide evidence that the mean differs from 50.

- One- sample chi - square test for a variance

What this test answers : "Is the population variance σ^2 equal to a specified value σ_0^2 ?"

- Assumptions : random sample $X_1, \dots, X_n$, population is Normal
- Hypotheses :
  - Two - sided : $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$
  - Upper - tailed : $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$
  - Lower - tailed : $H_0 : \sigma^2 \geq \sigma_0^2$ vs $H_1 : \sigma^2 < \sigma_0^2$
- Test statistic (under H_0) : Let $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ be the sample variance and $\nu = n - 1$ df Then

$$\chi_0^2 = \frac{(n-1) S^2}{\sigma_0^2} \sim \chi_\nu^2 \text{ if } H_0 \text{ is true}$$

- Decision rules at significance level α :
  - Using critical values :
    - Two - sided : reject H_0 if $\chi_0^2 \leq \chi^2_{\frac{\alpha}{2}, \nu}$ or $\chi_0^2 \geq \chi^2_{1-\frac{\alpha}{2}, \nu}$
    - Upper - tailed : reject if $\chi_0^2 \geq \chi^2_{1-\alpha, \nu}$
    - Lower - tailed : reject if $\chi_0^2 \leq \chi^2_{\alpha, \nu}$
  - Using the P- value :
    - Two - sided : $p = 2 \min\{P(\chi_\nu^2 \leq \chi_0^2), P(\chi_\nu^2 \geq \chi_0^2)\}$
    - Upper - tailed : $p = P(\chi_\nu^2 \geq \chi_0^2)$
    - Lower - tailed : $p = P(\chi_\nu^2 \leq \chi_0^2)$

    Reject H_0 if $p \leq \alpha$

Ex : You measure a Normal characteristic and want to check the spec variance $\sigma_0^2 = 1.0$.
Sample size $n = 10 \rightarrow \nu = 9$. From data, $s^2 = 1.44$.

Test $H_0 : \sigma^2 = 1.0$ vs $H_1 : \sigma^2 \neq 1.0$ at $\alpha = 0.05$.

1. Test statistic : $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \times 1.44}{1.00} = 12.96$

2. Critical values : From standard chi - squared tables,
$\chi^2_{0.025, 9} = 2.700 \quad \chi^2_{0.975, 9} = 19.023$

3. Decision : Acceptance region $[2.700, 19.023]$. Since 12.96 lies inside $\rightarrow$ fail to reject H_0.

At 5% level, the data do not show evidence that the process variance differs from 1.0.

- One - sample z - test for a population proportion p

- Assumptions : n independant, identical Bernoulli trials with success probability p.
Large - sample condition so the normal approximation is reasonable under H_0 : $np_0 \gtrsim 10$ and $n(1-p_0) \gtrsim 10$

- Hypotheses :
  - Two - sided : $H_0 : p = p_0$ vs $H_1 : p \neq p_0$
  - Upper - tailed : $H_0 : p \leq p_0$ vs $H_1 : p > p_0$
  - Lower - tailed : $H_0 : p \geq p_0$ vs $H_1 : p < p_0$

Let $\hat{p} = \frac{X}{n}$ be the sample proportion (with X successes)

- Test statistic (under H_0) : Use the standard error under the null :

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \overset{H_0}{\approx} \mathcal{N}(0,1)$$

- Decision rules at significance level α :
    - Using critical values :
        - Two - sided : reject H_0 if $|z_0| \geq z_{1-\frac{\alpha}{2}}$
        - Upper - tailed : reject if $z_0 \geq z_{1-\alpha}$
        - Lower - tailed : reject if $z_0 \leq z_\alpha$ (note $z_\alpha < 0$)
    - Using the P- value :
        - Two - sided : $p = 2[1 - \Phi(|z_0|)]$
        - Upper - tailed : $p = 1 - \Phi(z_0)$
        - Lower - tailed : $p = \Phi(z_0)$

    Reject H_0 if $p \leq \alpha$