Hypothesis testing

$$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F_\theta$$

H_0 versus H_1 about θ

- Four possibilities in total

		Decision	
		Do not reject H_0	Reject H_0
H_0	True	Correct Decision	Type I Error
	False	Type II Error	Correct Decision

- **Type I error:** reject H_0 when H_0 is true $\rightsquigarrow$ false positive
- **Type II error:** do not reject H_0 when H_0 is false $\rightsquigarrow$ false negative

Testing mean w/ known variance

$$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F,\ \mathbb{E}[X_1] = \mu,\ \text{Var}(X_i) = \sigma^2$$

$$\mathbb{P}_{H_0}(\text{reject } H_0) = \mathbb{P}_{H_0}(|\bar{X} - \mu_0| \geq C) = \mathbb{P}_{H_0}\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \geq C\frac{\sqrt{n}}{\sigma}\right) \approx 2\left(1 - \Phi\left(\frac{C\sqrt{n}}{\sigma}\right)\right)$$

$$\text{Setting } 2\left(1 - \Phi\left(\frac{C\sqrt{n}}{\sigma}\right)\right) = \alpha \Rightarrow \frac{C\sqrt{n}}{\sigma} = z_{\alpha/2} \Rightarrow C = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\text{Reject } H_0 \text{ when } |\bar{X} - \mu_0| \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Signals from the same star are independently received 20 times
- Sample mean = 11.6
- Significance level $\alpha = 0.05$
- Do we reject the H_0?

$$|11.6 - 10| = 1.6 \quad < \quad z_{0.025} \times 4/\sqrt{20} = 1.75$$

Fail to reject the null hypothesis at level 0.05!

Testing mean w/ known variance

Normal population or $n \geq 30$

- **Two-sided test**
  - $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0 \rightsquigarrow$ Reject H_0 when $|\bar{X} - \mu_0| \geq z_{\alpha/2}\sigma/\sqrt{n}$
- **One-sided test**
  - $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0 \rightsquigarrow$ Reject H_0 when $\bar{X} - \mu_0 \geq z_\alpha\sigma/\sqrt{n}$
  - $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0 \rightsquigarrow$ Reject H_0 when $\bar{X} - \mu_0 \leq -z_\alpha\sigma/\sqrt{n}$

Testing mean w/ unknown variance

$n \geq 30 \rightsquigarrow S \approx \sigma$

- **Two-sided test**
  - $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0 \rightsquigarrow$ Reject H_0 when $|\bar{X} - \mu_0| \geq z_{\alpha/2}S/\sqrt{n}$
- **One-sided test**
  - $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0 \rightsquigarrow$ Reject H_0 when $\bar{X} - \mu_0 \geq z_\alpha S/\sqrt{n}$
  - $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0 \rightsquigarrow$ Reject H_0 when $\bar{X} - \mu_0 \leq -z_\alpha S/\sqrt{n}$

Is the average number of chips in a bar greater than 1200?

- $\mu \rightsquigarrow$ the mean number of chips in a bag

$$H_0 : \mu \leq 1200 \text{ versus } H_1 : \mu > 1200$$

- Significance level $\alpha = 0.05$
- Rejection rule: reject H_0 when $\bar{X} - \mu_0 \geq z_\alpha S/\sqrt{n}$

$$\bar{X} - \mu_0 = 1261.57 - 1200 = 61.57 > 29.84 = z_{0.05}S/\sqrt{n}$$

Reject the null hypothesis at level 0.05!

Example: Chips Ahoy! (revisited)

- $\mu \rightsquigarrow$ the mean number of chips in a bag $H_0 : \mu \leq 1200$ versus $H_1 : \mu > 1200$
- n = 42
- Sample mean = 1261.57
- Sample standard deviation = 117.58
- $\alpha = 0.05$

$$t = \frac{1261.57 - 1200}{117.58/\sqrt{42}} = 3.39 \qquad p = \mathbb{P}(Z \geq 3.39) = 0.00035 \leq 0.05$$

Reject the null hypothesis at level 0.05!

- Among a clinic's patients having high cholesterol levels of at least 240 ml/dl of blood serum, volunteers were recruited to test a new drug designed to reduce blood cholesterol
- A group of 10 volunteers were given the drug for 60 days
- Change of blood cholesterol levels were recorded
- Average change was a decrease of 3.4
- Sample standard deviation is 12.1

- Test the hypothesis that the change in blood cholesterol was purely due to change
- $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$
- Assume normal population
- Significance level $\alpha = 0.05$

$$|\bar{X} - \mu_0| = 3.4 < 8.66 = t_{9,0.025}\frac{S}{\sqrt{10}}$$

Fail to reject the null hypothesis at level 0.05!

Test on variance in the normal population

$$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 \neq \sigma_0^2$$

Recall that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ $\Rightarrow$ Under H_0 $\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$

$$\text{Reject } H_0 \text{ if } \frac{(n-1)S^2}{\sigma_0^2} \geq \chi^2_{n-1,\alpha/2} \text{ or } \frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{n-1,1-\alpha/2}$$

- Cereal boxes are filled by a manufacturing company using a particular piece of equipment, which filled boxes according to a normal distribution with mean 368 grams and standard deviation 15 grams
- The piece of equipment fails and the manufacturing company orders a replacement.
- 25 cereal boxes produced by the new equipment are selected at random
- The sample standard deviation is computed to be 17.7 grams
- Test whether the standard deviation of the equipment has changed ($\alpha = 0.05$)

$$H_0 : \sigma^2 = 225 \text{ versus } H_1 : \sigma^2 \neq 225$$

$$\text{Reject } H_0 \text{ if } \frac{24 \times S^2}{225} \geq \chi^2_{24,0.025} \text{ or } \frac{24 \times S^2}{225} \leq \chi^2_{24,0.975}$$

$$\frac{24S^2}{225} = 33.418 \qquad \chi^2_{24,0.975} = 12.401 \qquad \chi^2_{24,0.025} = 39.364$$

Fail to reject H_0 at level 0.05

Table 7.6 Level α Tests on σ^2 (Test Statistic: $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$)

Testing Problem	Hypotheses	Reject H_0 if	P-value
Upper One-Sided	$H_0: \sigma^2 \le \sigma_0^2$ vs. $H_1: \sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{n-1,\alpha} \iff s^2 > \frac{\sigma_0^2 \chi^2_{n-1,\alpha}}{n-1}$	$P_U = P(\chi^2_{n-1} \ge \chi^2)$
Lower One-Sided	$H_0: \sigma^2 \ge \sigma_0^2$ vs. $H_1: \sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{n-1,1-\alpha} \iff s^2 < \frac{\sigma_0^2 \chi^2_{n-1,1-\alpha}}{n-1}$	$P_L = P(\chi^2_{n-1} \le \chi^2)$
Two-Sided	$H_0: \sigma^2 = \sigma_0^2$ vs. $H_1: \sigma^2 \ne \sigma_0^2$	$\chi^2 > \chi^2_{n-1,\alpha/2}$ or $\chi^2 < \chi^2_{n-1,1-\alpha/2} \iff s^2 > \frac{\sigma_0^2 \chi^2_{n-1,\alpha/2}}{n-1}$ or $s^2 < \frac{\sigma_0^2 \chi^2_{n-1,1-\alpha/2}}{n-1}$	$2\min\{P_U, P_L = 1 - P_U\}$

Duality between CI and HT: example

- $H_0: \mu = \mu_0$ versus $H_1: \mu \ne \mu_0$
- Suppose σ known

At level α, reject H_0 if $\dfrac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \ge z_{\alpha/2}$

$(1-\alpha)$-CI: $\left[\bar{X} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$

Reject $H_0 \iff \theta_0 \notin$ CI

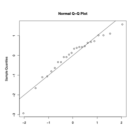
- In the final exam (total points 100):
  - Class 1: mean = 75, SD = 15
  - Class 2: mean = 70, SD = 12
- Is there a statistically significant difference at level $\alpha = 0.05$?

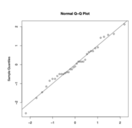
$H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \ne \mu_2$

$$|\bar{X} - \bar{Y}| = |75 - 70| = 5 \quad > \quad z_{\alpha/2}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = z_{0.025} \times \sqrt{\frac{15^2}{150} + \frac{12^2}{150}} = 3.074$$

We can therefore reject the H_0

	Perished	Survived
Mean	727.9	738
SD	23.54	19.84
Sample size	24	35





- Assume equal variance
- 95% CI for the average arm bone difference between the survival group and the perish group

$$S^2_{\text{pooled}} = \frac{23 \times 23.54^2 + 34 \times 19.84^2}{24 + 35 - 2} = 458.39$$

$$\left[738 - 727.9 - 2.00 \times \sqrt{458.39}\sqrt{1/24 + 1/35}, 738 - 727.9 + 2.00\sqrt{458.39}\sqrt{1/24 + 1/35}\right] = [-1.26, 21.46]$$

CI of the population mean

$X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$ ⟵ Independent ⟶ $Y_1, Y_2, \cdots, Y_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$

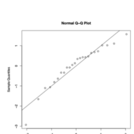
- Small sample!
- σ_1, σ_2 unknown but $\sigma_1 = \sigma_2 = \sigma$
- Two sided $(1-\alpha)$ CI for $\mu_1 - \mu_2$

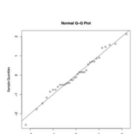
$$\left[\bar{X} - \bar{Y} - t_{n+m-2,\alpha/2} S\sqrt{\frac{1}{n} + \frac{1}{m}}, \bar{X} - \bar{Y} + t_{n+m-2,\alpha/2} S\sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

- Test $H_0: \mu_1 - \mu_2 = \delta$ versus $\mu_1 - \mu_2 \ne \delta_0$

Reject H_0 when $\left|\bar{X} - \bar{Y} - \delta_0\right| \ge t_{n+m-2,\alpha/2} S\sqrt{\frac{1}{n} + \frac{1}{m}}$

	Perished	Survived
Mean	727.9	738
SD	23.54	19.84
Sample size	24	35





- **Do not** assume equal variance
- 95% CI for the average arm bone difference between the survival group and the perish group

$w_1 = 23.54^2/24 = 22.87$

$w_2 = 19.85^2/35 = 11.25$

$$\nu = \frac{(w_1 + w_2)^2}{w_1^2/(n-1) + w_2^2/(m-1)} \approx 43$$

CI=[-1.72,21.92]

Previously in the equal variance case: [-1.27, 21.47]

Comparing variances

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1,m-1}$$

- Testing $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_1: \sigma_1^2 \ne \sigma_2^2$

$$\frac{S_1^2}{S_2^2} > f_{n-1,m-1,\alpha/2} \text{ or } \frac{S_1^2}{S_2^2} < f_{n-1,m-1,1-\alpha/2}$$

Example: coffee and blood flow during exercise

Namdar et al (2006). Caffeine decreases exercise-induced myocardial flow reserve. *Journal of the American College of Cardiology 47, 405-410.*

- Doctors studying healthy men measured myocardial blood flow (MBF) during bicycle exercise after giving the subjects a placebo or a dose of 200 mg of caffeine that was equivalent to drinking two cups of coffee
- 8 subjects ⤳ each tested twice
- 4 of them randomly selected to receive caffeine in the first test and placebo in the second
- The other 4 received placebo first and caffeine second
- 24-hour gap between the two sets (washout period)

XX

Comparison of means with matched pairs

$$D_i \overset{iid}{\sim} \mathcal{N}(\mu_1 - \mu_2, \sigma_D^2)$$

- Sample variance of the differences S_D
- Pivotal statistics

$$\frac{\bar{D} - (\mu_1 - \mu_2)}{S_D/\sqrt{n}} \sim t_{n-1}$$

- Test $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \ne \mu_2$

Observed test statistic $t = \dfrac{\bar{D}}{S_D/\sqrt{n}}$, p-value = $\mathbb{P}(|t_{n-1}| \ge |t|)$

Example: testing uniformity of random digits (cont'd)

Digit	0	1	2	3	4	5	6	7	8	9
Observed Count	12	7	12	7	13	13	7	13	6	10

- p_i: the probability of getting i
- Test $H_0: p_0 = p_1 = \ldots = p_{10} = \dfrac{1}{10}$

$$\chi^2 = \frac{(12-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(3-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(13-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(10-10)^2}{10} = 7.8 < 16.92 = \chi^2_{9,0.05}$$

Fail to reject H_0

Test of Law of independent assortment

$$H_0: p_1 = 9/16,\ p_2 = 3/16,\ p_3 = 3/16,\ p_4 = 1/16$$

	Yellow/Round	Green/Round	Yellow/Wrinkled	Green/Wrinkled
Observed	315	108	102	31
Expected	$\frac{9}{16} \times 556 = 312.75$	$\frac{3}{16} \times 556 = 104.25$	$\frac{3}{16} \times 556 = 104.25$	$\frac{1}{16} \times 556 = 34.75$

$$\chi^2 \approx 0.604 < 7.81 = \chi^2_{3,0.95}$$

Fail to reject H_0

	Married	Single	Widowed or Divorced	Total
Severe	22	16	19	57
Normal	33	29	14	76
Mild	14	9	3	26
Total	69	54	36	159

	Married	Single	Widowed or Divorced
Severe	$\frac{57\times69}{159} = 24.7$	$\frac{57\times54}{159} = 19.4$	$\frac{57\times36}{159} = 12.9$
Normal	$\frac{76\times69}{159} = 33.6$	$\frac{76\times54}{159} = 25.8$	$\frac{76\times36}{159} = 17.2$
Mild	$\frac{26\times69}{159} = 11.3$	$\frac{26\times54}{159} = 8.8$	$\frac{26\times36}{159} = 5.9$

$$\chi^2 \approx 6.85 < 9.49 = \chi^2_{4,0.95}$$

Fail to reject H_0