Statistics for Astronomers: Lecture 13, 2020.12.02

Prof. Sundar Srinivasan

IRyA/UNAM

Review

Hypothesis testing.

- Null hypothesis, simple and composite hypotheses. One/two-tailed hypotheses.
- Type I and II errors, p-value, statistical power.
- Likelihood-ratio test.
- One-sample Z- and t-tests.

2-sample tests: independent & dependent/paired samples

Independent samples $\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$.

Dependent samples: $\{x_{1,i}\}$ and $\{x_{2,i}\}$, $i = 1 \cdots N$, such that $x_{1,i}$ related to $x_{2,i}$.

$$\mathrm{Var}[\overline{x_1} - \overline{x_2}]$$

2-sample tests: independent & dependent/paired samples

Independent samples $\text{Var}[\overline{x_1} - \overline{x_2}] = \text{Var}[\overline{x_1}] + \text{Var}[\overline{x_2}]$.

Dependent samples: $\{x_{1,i}\}$ and $\{x_{2,i}\}$, $i = 1 \cdots N$, such that $x_{1,i}$ related to $x_{2,i}$.

$$\text{Var}[\overline{x_1} - \overline{x_2}] = \text{Var}[\overline{x_1}] + \text{Var}[\overline{x_2}] - 2\rho\sqrt{\text{Var}[\overline{x_1}]\text{Var}[\overline{x_2}]}.$$

2-sample tests: independent & dependent/paired samples

Independent samples $\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$.

Dependent samples: $\{x_{1,i}\}$ and $\{x_{2,i}\}$, $i = 1 \cdots N$, such that $x_{1,i}$ related to $x_{2,i}$.

$$\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}] - 2\rho\sqrt{\mathrm{Var}[\overline{x_1}]\mathrm{Var}[\overline{x_2}]}.$$

If $\rho > 0, \mathrm{Var}[\overline{x_1} - \overline{x_2}] < \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$ and vice versa.

Also called paired/matched/correlated samples.

2-sample tests: independent & dependent/paired samples

Independent samples $\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$.

Dependent samples: $\{x_{1,i}\}$ and $\{x_{2,i}\}$, $i = 1 \cdots N$, such that $x_{1,i}$ related to $x_{2,i}$.

$$\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}] - 2\rho\sqrt{\mathrm{Var}[\overline{x_1}]\mathrm{Var}[\overline{x_2}]}.$$

If $\rho > 0, \mathrm{Var}[\overline{x_1} - \overline{x_2}] < \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$ and vice versa.

Also called paired/matched/correlated samples.

Example: Flux in the pixels of an image before and after background subtraction.

$S_{2,i} = S_{1,i} - B_i$ strong correlation, typically $\rho \approx 1$.

H_0: The mean flux per pixel is the same after background subtraction.

One way to reduce overall variance is to pair samples ("beating $\sqrt{N}$"; see Barlow).

2-sample tests: independent & dependent/paired samples

Independent samples $\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$.

Dependent samples: $\{x_{1,i}\}$ and $\{x_{2,i}\}$, $i = 1 \cdots N$, such that $x_{1,i}$ related to $x_{2,i}$.

$$\mathrm{Var}[\overline{x_1} - \overline{x_2}] = \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}] - 2\rho\sqrt{\mathrm{Var}[\overline{x_1}]\mathrm{Var}[\overline{x_2}]}.$$

If $\rho > 0, \mathrm{Var}[\overline{x_1} - \overline{x_2}] < \mathrm{Var}[\overline{x_1}] + \mathrm{Var}[\overline{x_2}]$ and vice versa.

Also called paired/matched/correlated samples.

Example: Flux in the pixels of an image before and after background subtraction.
$S_{2,i} = S_{1,i} - B_i$ strong correlation, typically $\rho \approx 1$.
H_0: The mean flux per pixel is the same after background subtraction.
One way to reduce overall variance is to pair samples ("beating $\sqrt{N}$"; see Barlow).

Since we typically compute statistics in terms of the variance (*e.g.*, by standardisation), the behaviour of the statistic changes for dependent samples.

As $\rho \uparrow, \mathrm{Var}[\text{difference between means}] \downarrow$
For a fixed threshold/critical value, $P(\text{reject } H_0 | H_0 \text{ true}) \downarrow$, Type I error $\downarrow$, power $\uparrow$.

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}$; $\mathbb{E}[\overline{y}] =$

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}$; $\mathbb{E}[\overline{y}] = \mu_1 - \mu_2$; $\mathrm{Var}[\overline{y}] =$

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

$$\text{Define } Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}; \quad \mathbb{E}[\overline{y}] = \mu_1 - \mu_2; \quad \mathrm{Var}[\overline{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - \cancel{2\rho\frac{\sigma_1\sigma_2}{\sqrt{N_1 N_2}}} \text{ (independence)}$$

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}$; $\quad \mathbb{E}[\overline{y}] = \mu_1 - \mu_2$; $\quad \mathrm{Var}[\overline{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - 2\rho \frac{\sigma_1\sigma_2}{\sqrt{N_1 N_2}}$ (independence)

Standardise: $Z \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\mathrm{Var}[\overline{y}]}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$ under H_0: $\mu_1 = \mu_2$.

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}$; $\quad \mathbb{E}[\overline{y}] = \mu_1 - \mu_2$; $\quad \mathrm{Var}[\overline{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - 2\rho \frac{\sigma_1 \sigma_2}{\sqrt{N_1 N_2}}$ (independence)

Standardise: $Z \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\mathrm{Var}[\overline{y}]}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$ under H_0: $\mu_1 = \mu_2$.

Example:

Sarah selects 50 AGN from a famous dataset of Type-I AGN and 45 AGN from her own dataset.
The population standard deviations of the SFRs of the samples are 1.8 and 0.95 $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ respectively.
Sarah finds sample means of 2.2 and 3.2 $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ respectively. At the 95% confidence level, does her dataset consist of AGN with systematically higher SFRs than those of the famous dataset?

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}$; $\quad \mathbb{E}[\overline{y}] = \mu_1 - \mu_2$; $\quad \mathrm{Var}[\overline{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - \cancel{2\rho \frac{\sigma_1\sigma_2}{\sqrt{N_1 N_2}}}$ (independence)

Standardise: $$Z \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\mathrm{Var}[\overline{y}]}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$ under H_0: $\mu_1 = \mu_2$.

Example:

Sarah selects 50 AGN from a famous dataset of Type-I AGN and 45 AGN from her own dataset.
The population standard deviations of the SFRs of the samples are 1.8 and 0.95 $M_\odot\ \mathrm{yr}^{-1}$ respectively.
Sarah finds sample means of 2.2 and 3.2 $M_\odot\ \mathrm{yr}^{-1}$ respectively. At the 95% confidence level, does her dataset consist of AGN with systematically higher SFRs than those of the famous dataset?

H_0: $\mu_2 = \mu_1$. H_A: $\mu_2 > \mu_1$ (right-tailed test).

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \bar{y} = \bar{x_1} - \bar{x_2}$; $\mathbb{E}[\bar{y}] = \mu_1 - \mu_2$; $\mathrm{Var}[\bar{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - 2\rho\frac{\sigma_1\sigma_2}{\sqrt{N_1 N_2}}$ (independence)

Standardise: $Z \equiv \frac{\bar{y} - \mathbb{E}[\bar{y}]}{\sqrt{\mathrm{Var}[\bar{y}]}} = \frac{\bar{x_1} - \bar{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$ under H_0: $\mu_1 = \mu_2$.

Example:

Sarah selects 50 AGN from a famous dataset of Type-I AGN and 45 AGN from her own dataset.
The population standard deviations of the SFRs of the samples are 1.8 and 0.95 $M_\odot$ yr^{-1} respectively.
Sarah finds sample means of 2.2 and 3.2 $M_\odot$ yr^{-1} respectively. At the 95% confidence level, does her dataset consist of AGN with systematically higher SFRs than those of the famous dataset?

H_0: $\mu_2 = \mu_1$. H_A: $\mu_2 > \mu_1$ (right-tailed test).

$\bar{x_i} \sim \mathcal{N}(\mu_i, \sigma_i^2/N_i)$, with $i = 1, 2 \implies x_2 - x_1 \sim \mathcal{N}(\mu_2 - \mu_1, \sigma_2^2/N_1 + \sigma_1^2/N_2)$.

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

Define $Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}$; $\mathbb{E}[\overline{y}] = \mu_1 - \mu_2$; $\mathrm{Var}[\overline{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - 2\rho\frac{\sigma_1\sigma_2}{\sqrt{N_1 N_2}}$ (independence)

Standardise: $Z \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\mathrm{Var}[\overline{y}]}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$ under H_0: $\mu_1 = \mu_2$.

Example:

Sarah selects 50 AGN from a famous dataset of Type-I AGN and 45 AGN from her own dataset.
The population standard deviations of the SFRs of the samples are 1.8 and 0.95 $M_\odot$ yr^{-1} respectively.
Sarah finds sample means of 2.2 and 3.2 $M_\odot$ yr^{-1} respectively. At the 95% confidence level, does her dataset consist of AGN with systematically higher SFRs than those of the famous dataset?

H_0: $\mu_2 = \mu_1$. H_A: $\mu_2 > \mu_1$ (right-tailed test).

$\overline{x_i} \sim \mathcal{N}(\mu_i, \sigma_i^2/N_i)$, with $i = 1, 2 \implies x_2 - x_1 \sim \mathcal{N}(\mu_2 - \mu_1, \sigma_2^2/N_1 + \sigma_1^2/N_2)$.

$Z \equiv \frac{\overline{x_2} - \overline{x_1} - (\mu_2 - \mu_1)}{\sqrt{\sigma_2^2/N_1 + \sigma_1^2/N_2}} = \frac{\overline{x_2} - \overline{x_1}}{\sqrt{\sigma_2^2/N_1 + \sigma_1^2/N_2}}$ (because $\mu_2 = \mu_1$ under H_0) ≈ 3.43.

Two-sample Z-test

Independent samples $\{x_{1,i}\}$ (N_1 points), $\{x_{2,i}\}$ (N_2 points) with $X_j \sim \mathcal{N}(\mu_j, \sigma_j^2)$, $j = 1, 2$.
Question: is $\mu_1 = \mu_2$? Convert to a one-sample problem:

$$\text{Define } Y = x_1 - x_2 \implies \overline{y} = \overline{x_1} - \overline{x_2}; \quad \mathbb{E}[\overline{y}] = \mu_1 - \mu_2; \quad \mathrm{Var}[\overline{y}] = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} - 2\rho\frac{\sigma_1\sigma_2}{\sqrt{N_1 N_2}} \text{ (independence)}$$

$$\text{Standardise: } Z \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\mathrm{Var}[\overline{y}]}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \text{ under } H_0\text{: } \mu_1 = \mu_2.$$

Example:

Sarah selects 50 AGN from a famous dataset of Type-I AGN and 45 AGN from her own dataset.
The population standard deviations of the SFRs of the samples are 1.8 and 0.95 $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ respectively.
Sarah finds sample means of 2.2 and 3.2 $\mathrm{M}_\odot\ \mathrm{yr}^{-1}$ respectively. At the 95% confidence level, does her dataset consist of AGN with systematically higher SFRs than those of the famous dataset?

H_0: $\mu_2 = \mu_1$. H_A: $\mu_2 > \mu_1$ (right-tailed test).

$\overline{x_i} \sim \mathcal{N}(\mu_i, \sigma_i^2/N_i)$, with $i = 1, 2 \implies x_2 - x_1 \sim \mathcal{N}(\mu_2 - \mu_1, \sigma_2^2/N_1 + \sigma_1^2/N_2)$.

$$Z \equiv \frac{\overline{x_2} - \overline{x_1} - (\mu_2 - \mu_1)}{\sqrt{\sigma_2^2/N_1 + \sigma_1^2/N_2}} = \frac{\overline{x_2} - \overline{x_1}}{\sqrt{\sigma_2^2/N_1 + \sigma_1^2/N_2}} \text{ (because } \mu_2 = \mu_1 \text{ under } H_0) \approx 3.43.$$

p-value: $P(Z > 3.43) \approx 0.0003 < \alpha = 0.05$, therefore H_0 can be rejected at 5% significance.

Two-sample t-tests: independent samples

If σ_1, σ_2 unknown and $\sigma_1 = \sigma_2$, can use "regular" t-test if $N_1 \approx N_2$.

If $N_1 \approx N_2$, can also use "regular" test when $\sigma_1 \neq \sigma_2$.

If σ_1, σ_2 unknown and $\sigma_1 \neq \sigma_2$ or $N_1 \neq N_2$, use Welch's t-test.

If we don't know σ_1, σ_2, how the hell can we know if they are (un)equal?! – F-test.

Two-sample t-tests: independent samples

If σ_1, σ_2 unknown and $\sigma_1 = \sigma_2$, can use "regular" t-test if $N_1 \approx N_2$.

If $N_1 \approx N_2$, can also use "regular" test when $\sigma_1 \neq \sigma_2$.

If σ_1, σ_2 unknown and $\sigma_1 \neq \sigma_2$ or $N_1 \neq N_2$, use Welch's t-test.

If we don't know σ_1, σ_2, how the hell can we know if they are (un)equal?! – F-test.

Once again, define $\overline{y}$ as the difference in means $\overline{x_1} - \overline{x_2}$.
For the two-sample Z-test, we standardised $\overline{y}$.
For the two-sample t-test, we studentise instead:

Two-sample t-tests: independent samples

If σ_1, σ_2 unknown and $\sigma_1 = \sigma_2$, can use "regular" t-test if $N_1 \approx N_2$.

If $N_1 \approx N_2$, can also use "regular" test when $\sigma_1 \neq \sigma_2$.

If σ_1, σ_2 unknown and $\sigma_1 \neq \sigma_2$ or $N_1 \neq N_2$, use Welch's t-test.

If we don't know σ_1, σ_2, how the hell can we know if they are (un)equal?! – F-test.

Once again, define $\overline{y}$ as the difference in means $\overline{x_1} - \overline{x_2}$.
For the two-sample Z-test, we standardised $\overline{y}$.
For the two-sample t-test, we studentise instead:

$$t \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\mathrm{Var}[\overline{y}]}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \text{ under } H_0\text{: } \mu_1 = \mu_2.$$

Two-sample t-tests: independent samples

If σ_1, σ_2 unknown and $\sigma_1 = \sigma_2$, can use "regular" t-test if $N_1 \approx N_2$.

If $N_1 \approx N_2$, can also use "regular" test when $\sigma_1 \neq \sigma_2$.

If σ_1, σ_2 unknown and $\sigma_1 \neq \sigma_2$ or $N_1 \neq N_2$, use Welch's t-test.

If we don't know σ_1, σ_2, how the hell can we know if they are (un)equal?! – F-test.

Once again, define $\overline{y}$ as the difference in means $\overline{x_1} - \overline{x_2}$.
For the two-sample Z-test, we standardised $\overline{y}$.
For the two-sample t-test, we studentise instead:

$$t \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\widehat{\mathrm{Var}[\overline{y}]}}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \text{ under } H_0\text{: } \mu_1 = \mu_2.$$

s_1 and s_2 are the (unbiased) sample standard deviations for $\{x_{1,i}\}$ and $\{x_{2,i}\}$ respectively.

This t-statistic has $\#dof = \nu = (N_1 - 1) + (N_2 - 1)$ (2 means computed from pooled data).

Two-sample t-tests: independent samples

If σ_1, σ_2 unknown and $\sigma_1 = \sigma_2$, can use "regular" t-test if $N_1 \approx N_2$.

If $N_1 \approx N_2$, can also use "regular" test when $\sigma_1 \neq \sigma_2$.

If σ_1, σ_2 unknown and $\sigma_1 \neq \sigma_2$ or $N_1 \neq N_2$, use Welch's t-test.

If we don't know σ_1, σ_2, how the hell can we know if they are (un)equal?! – F-test.

Once again, define $\overline{y}$ as the difference in means $\overline{x_1} - \overline{x_2}$.
For the two-sample Z-test, we standardised $\overline{y}$.
For the two-sample t-test, we studentise instead:

$$t \equiv \frac{\overline{y} - \mathbb{E}[\overline{y}]}{\sqrt{\widehat{\mathrm{Var}[\overline{y}]}}} = \frac{\overline{x_1} - \overline{x_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \text{ under } H_0\text{: } \mu_1 = \mu_2.$$

s_1 and s_2 are the (unbiased) sample standard deviations for $\{x_{1,i}\}$ and $\{x_{2,i}\}$ respectively.

This t-statistic has $\#dof = \nu = (N_1 - 1) + (N_2 - 1)$ (2 means computed from pooled data).

Implementation: `scipy.stats.ttest_ind`, more versatile than demonstrated here.

Two-sample t-tests: dependent/paired samples

$N_1 = N_2 = N$ means we can connect the i^{th} elements of the two samples.

Once again, define $\overline{y} = \overline{x_1} - \overline{x_2}$, but now use the fact that the samples are paired:

Two-sample t-tests: dependent/paired samples

$N_1 = N_2 = N$ means we can connect the $i^{\rm th}$ elements of the two samples.

Once again, define $\overline{y} = \overline{x_1} - \overline{x_2}$, but now use the fact that the samples are paired:

$$\overline{y} = \overline{x_1} - \overline{x_2} = \frac{1}{N}\left(\sum_{i=1}^{N} x_{1,i} - \sum_{i=1}^{N} x_{2,i}\right) = \frac{1}{N}\sum_{i=1}^{N}(x_{1,i} - x_{2,i}) \equiv \frac{1}{N}\sum_{i=1}^{N} y_i.$$

(we can do this since $x_{1,i}$ is correlated with $x_{2,i}$)

Two-sample t-tests: dependent/paired samples

$N_1 = N_2 = N$ means we can connect the $i^{\rm th}$ elements of the two samples.

Once again, define $\overline{y} = \overline{x_1} - \overline{x_2}$, but now use the fact that the samples are paired:

$$\overline{y} = \overline{x_1} - \overline{x_2} = \frac{1}{N}\left(\sum_{i=1}^{N} x_{1,i} - \sum_{i=1}^{N} x_{2,i}\right) = \frac{1}{N}\sum_{i=1}^{N}(x_{1,i} - x_{2,i}) \equiv \frac{1}{N}\sum_{i=1}^{N} y_i.$$

(we can do this since $x_{1,i}$ is correlated with $x_{2,i}$)

Unbiased sample std $s_y \equiv \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\overline{y}^2\right]}$

(Only one mean, $\overline{y}$, is computed from the data $\{y_i\}$)

Two-sample t-tests: dependent/paired samples

$N_1 = N_2 = N$ means we can connect the i^{th} elements of the two samples.

Once again, define $\overline{y} = \overline{x_1} - \overline{x_2}$, but now use the fact that the samples are paired:

$$\overline{y} = \overline{x_1} - \overline{x_2} = \frac{1}{N}\left(\sum_{i=1}^{N} x_{1,i} - \sum_{i=1}^{N} x_{2,i}\right) = \frac{1}{N}\sum_{i=1}^{N}(x_{1,i} - x_{2,i}) \equiv \frac{1}{N}\sum_{i=1}^{N} y_i.$$

(we can do this since $x_{1,i}$ is correlated with $x_{2,i}$)

Unbiased sample std $s_y \equiv \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\overline{y}^2\right]}$

(Only one mean, $\overline{y}$, is computed from the data $\{y_i\}$)

Studentise $\overline{y}$: $t = \frac{\overline{y} - 0}{s_y/\sqrt{N}} = \frac{\overline{x_1} - \overline{x_2}}{s_y/\sqrt{N}}$ (if H_0 is true).

This t-statistic has $\#dof = \nu = N - 1$ (N data pairs, one mean computed).

Two-sample t-tests: dependent/paired samples

$N_1 = N_2 = N$ means we can connect the i^{th} elements of the two samples.

Once again, define $\overline{y} = \overline{x_1} - \overline{x_2}$, but now use the fact that the samples are paired:

$$\overline{y} = \overline{x_1} - \overline{x_2} = \frac{1}{N}\left(\sum_{i=1}^{N} x_{1,i} - \sum_{i=1}^{N} x_{2,i}\right) = \frac{1}{N}\sum_{i=1}^{N}(x_{1,i} - x_{2,i}) \equiv \frac{1}{N}\sum_{i=1}^{N} y_i.$$

(we can do this since $x_{1,i}$ is correlated with $x_{2,i}$)

Unbiased sample std $s_y \equiv \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{y})^2} = \sqrt{\frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\overline{y}^2\right]}$

(Only one mean, $\overline{y}$, is computed from the data $\{y_i\}$)

Studentise $\overline{y}$: $t = \frac{\overline{y} - 0}{s_y/\sqrt{N}} = \frac{\overline{x_1} - \overline{x_2}}{s_y/\sqrt{N}}$ (if H_0 is true).

This t-statistic has $\#dof = \nu = N - 1$ (N data pairs, one mean computed).

Implementation: `scipy.stats.ttest_rel`, more versatile than demonstrated here.

Trying out t-tests

1. Download this Jupyter notebook.
2. Navigate to Colaboratory.
3. Sign in
4. Click on "Upload" and upload the notebook you downloaded in step 1.

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

Under H_0, the ratio $F \equiv \dfrac{S_2^2}{S_1^2} \geq 1$ is a ratio of two reduced χ^2 variables of the form $\dfrac{\chi_\nu^2}{\nu}$.

This ratio has the F-distribution with degrees of freedom $(\nu_2 = N_2 - 1, \nu_1 = N_1 - 1)$.

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

Under H_0, the ratio $F \equiv \dfrac{S_2^2}{S_1^2} \geq 1$ is a ratio of two reduced χ^2 variables of the form $\dfrac{\chi_\nu^2}{\nu}$.

This ratio has the F-distribution with degrees of freedom $(\nu_2 = N_2 - 1, \nu_1 = N_1 - 1)$.

Asymptotic behaviour: for large N_1, N_2,

$$Z = \frac{1}{2}\ln F \text{ is approximately distributed as } \mathcal{N}\left(\tfrac{1}{2}(1/\nu_1 - 1/\nu_2), \tfrac{1}{2}(1/\nu_1 + 1/\nu_2)\right).$$

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

Under H_0, the ratio $F \equiv \dfrac{S_2^2}{S_1^2} \geq 1$ is a ratio of two reduced χ^2 variables of the form $\dfrac{\chi_\nu^2}{\nu}$.

This ratio has the F-distribution with degrees of freedom $(\nu_2 = N_2 - 1, \nu_1 = N_1 - 1)$.

Asymptotic behaviour: for large N_1, N_2,

$$Z = \frac{1}{2}\ln F \text{ is approximately distributed as } \mathcal{N}\left(\tfrac{1}{2}(1/\nu_1 - 1/\nu_2), \tfrac{1}{2}(1/\nu_1 + 1/\nu_2)\right).$$

Example (Barlow):

$$N_1 = 12, S_1^2 = 10.9, N_2 = 7, S_2^2 = 6.5 \Longrightarrow F = \frac{S_1^2}{S_2^2} = 1.68.$$

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

Under H_0, the ratio $F \equiv \dfrac{S_2^2}{S_1^2} \geq 1$ is a ratio of two reduced χ^2 variables of the form $\dfrac{\chi^2_\nu}{\nu}$.

This ratio has the F-distribution with degrees of freedom $(\nu_2 = N_2 - 1, \nu_1 = N_1 - 1)$.

Asymptotic behaviour: for large N_1, N_2,

$$Z = \frac{1}{2}\ln F \text{ is approximately distributed as } \mathcal{N}\left(\tfrac{1}{2}(1/\nu_1 - 1/\nu_2), \tfrac{1}{2}(1/\nu_1 + 1/\nu_2)\right).$$

Example (Barlow):

$$N_1 = 12, S_1^2 = 10.9, N_2 = 7, S_2^2 = 6.5 \Longrightarrow F = \frac{S_1^2}{S_2^2} = 1.68.$$

p-value $= P(F > 1.68) = 1 - P(F \leq 1.68)$

$= 1-$ `scipy.stats.f.cdf(1.68, 12-1, 7-1)`

$\approx 0.27 > \alpha = 0.05$.

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

Under H_0, the ratio $F \equiv \dfrac{S_2^2}{S_1^2} \geq 1$ is a ratio of two reduced χ^2 variables of the form $\dfrac{\chi^2_\nu}{\nu}$.

This ratio has the F-distribution with degrees of freedom $(\nu_2 = N_2 - 1, \nu_1 = N_1 - 1)$.

Asymptotic behaviour: for large N_1, N_2,

$$Z = \frac{1}{2}\ln F \text{ is approximately distributed as } \mathcal{N}\left(\tfrac{1}{2}(1/\nu_1 - 1/\nu_2), \tfrac{1}{2}(1/\nu_1 + 1/\nu_2)\right).$$

Example (Barlow):

$$N_1 = 12, S_1^2 = 10.9, N_2 = 7, S_2^2 = 6.5 \Longrightarrow F = \frac{S_1^2}{S_2^2} = 1.68.$$

p-value $= P(F > 1.68) = 1 - P(F \leq 1.68)$
$= 1-$ `scipy.stats.f.cdf(1.68, 12-1, 7-1)`
$\approx 0.27 > \alpha = 0.05$.

No statistical evidence for difference in the variances.
Safe to use t-test on these data assuming that $\sigma_1 = \sigma_2$.

See documentation for `scipy.stats.f` and Section 4.7.6 in the AstroML book.

F-test

Recall: If $X \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance $S^2 \sim \dfrac{\sigma^2}{N-1}\chi^2(N-1)$.

Consider two samples with sample variances $S_1^2 \leq S_2^2$. H_0: $\sigma_1 = \sigma_2 = \sigma$.

Under H_0, the ratio $F \equiv \dfrac{S_2^2}{S_1^2} \geq 1$ is a ratio of two reduced χ^2 variables of the form $\dfrac{\chi^2_\nu}{\nu}$.

This ratio has the F-distribution with degrees of freedom $(\nu_2 = N_2 - 1, \nu_1 = N_1 - 1)$.

Asymptotic behaviour: for large N_1, N_2,

$$Z = \frac{1}{2}\ln F \text{ is approximately distributed as } \mathcal{N}\left(\tfrac{1}{2}(1/\nu_1 - 1/\nu_2), \tfrac{1}{2}(1/\nu_1 + 1/\nu_2)\right).$$

Example (Barlow):

$N_1 = 12, S_1^2 = 10.9, N_2 = 7, S_2^2 = 6.5 \implies F = \dfrac{S_1^2}{S_2^2} = 1.68.$

p-value $= P(F > 1.68) = 1 - P(F \leq 1.68)$
$= 1-$ `scipy.stats.f.cdf(1.68, 12-1, 7-1)`
$\approx 0.27 > \alpha = 0.05$.

No statistical evidence for difference in the variances.
Safe to use t-test on these data assuming that $\sigma_1 = \sigma_2$.

Example:
Given $N_1 = 10$, $N_2 = 7$, and $S_2^2 = (1 + \lambda)\, S_1^2$; $(\lambda > 0)$,
find λ such that H_0 is rejected with 99.5% confidence.

See documentation for `scipy.stats.f` and Section 4.7.6 in the AstroML book.