# 🔭 Assignment: Measuring Cosmological Parameters Using Type Ia Supernovae

In this assignment, you'll analyze observational data from the Pantheon+SH0ES dataset of Type Ia supernovae to measure the Hubble constant $H_0$ and estimate the age of the universe. You will:

- Plot the Hubble diagram (distance modulus vs. redshift)
- Fit a cosmological model to derive $H_0$ and $\Omega_m$
- Estimate the age of the universe
- Analyze residuals to assess the model
- Explore the effect of fixing $\Omega_m$
- Compare low-z and high-z results

Let's get started!

## 📦 Getting Started: Setup and Libraries

Before we dive into the analysis, we need to import the necessary Python libraries:

- `numpy`, `pandas` — for numerical operations and data handling
- `matplotlib` — for plotting graphs
- `scipy.optimize.curve_fit` and `scipy.integrate.quad` — for fitting cosmological models and integrating equations
- `astropy.constants` and `astropy.units` — for physical constants and unit conversions

Make sure these libraries are installed in your environment. If not, you can install them using:

```
pip install numpy pandas matplotlib scipy astropy
```

In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
from scipy.integrate import quad
from astropy.constants import c
from astropy import units as u
```

## 📂 Load the Pantheon+SH0ES Dataset

We now load the observational supernova data from the Pantheon+SH0ES sample. This dataset includes calibrated distance moduli $\mu$, redshifts corrected for various effects, and uncertainties.

## Instructions:

- Make sure the data file is downloaded from Pantheon dataset and available locally.
- We use `delim_whitespace=True` because the file is space-delimited rather than comma-separated.
- Commented rows (starting with `#`) are automatically skipped.

We will extract:

- `zHD` : Hubble diagram redshift
- `MU_SH0ES` : Distance modulus using SH0ES calibration
- `MU_SH0ES_ERR_DIAG` : Associated uncertainty

More detailed column names and the meanings can be referred here:

> Finally, we include a combined file of all the fitted parameters for each SN, before and after light-curve cuts are applied. This is in the format of a .FITRES file and has all the meta-information listed above along with the fitted SALT2 parameters. We show a screenshot of the release in Figure 7. Here, we give brief descriptions of each column. **CID** – name of SN. **CIDint** – counter of SNe in the sample. **IDSURVEY** – ID of the survey. **TYPE** – whether SN Ia or not – all SNe in this sample are SNe Ia. **FIELD** – if observed in a particular field. **CUTFLAG_SNANA** – any bits in light-curve fit flagged. **ERRFLAG_FIT** – flag in fit. **zHEL** – heliocentric redshift. **zHELERR** – heliocentric redshift error. **zCMB** – CMB redshift. **zCMBERR** – CMB redshift error. **zHD** – Hubble Diagram redshift. **zHDERR** – Hubble Diagram redshift error. **VPEC** – peculiar velocity. **VPECERR** – peculiar-velocity error. **MWEBV** – MW extinction. **HOST_LOGMASS** – mass of host. **HOST_LOGMASS_ERR** – error in mass of host. **HOST_sSFR** – sSFR of host. **HOST_sSFR_ERR** – error in sSFR of host. **PKMJDINI** – initial guess for PKMJD. **SNRMAX1** – First highest signal-to-noise ratio (SNR) of light curve. **SNRMAX2** – Second highest SNR of light curve. **SNRMAX3** – Third highest SNR of light curve. **PKMJD** – Fitted PKMJD. **PKMJDERR** –

In [2]:
```python
# Local file path
file_path = r"D:\Academics & Jobs\ISA Summer School\2. Hubble Parameter\Pantheon

# Load the file
df = pd.read_csv(file_path, delim_whitespace=True, comment="#")

# See structure
df.head()
```

Out[2]:

| | CID | IDSURVEY | zHD | zHDERR | zCMB | zCMBERR | zHEL | zHELERR |
|---|---|---|---|---|---|---|---|---|
| **0** | 2011fe | 51 | 0.00122 | 0.00084 | 0.00122 | 0.00002 | 0.00082 | 0.00002 |
| **1** | 2011fe | 56 | 0.00122 | 0.00084 | 0.00122 | 0.00002 | 0.00082 | 0.00002 |
| **2** | 2012cg | 51 | 0.00256 | 0.00084 | 0.00256 | 0.00002 | 0.00144 | 0.00002 |
| **3** | 2012cg | 56 | 0.00256 | 0.00084 | 0.00256 | 0.00002 | 0.00144 | 0.00002 |
| **4** | 1994DRichmond | 50 | 0.00299 | 0.00084 | 0.00299 | 0.00004 | 0.00187 | 0.00004 |

5 rows × 47 columns

## Preview Dataset Columns

Before diving into the analysis, let's take a quick look at the column names in the dataset. This helps us verify the data loaded correctly and identify the relevant columns we'll use for cosmological modeling.

## 🧹 Clean and Extract Relevant Data

To ensure reliable fitting, we remove any rows that have missing values in key columns:

- `zHD` : redshift for the Hubble diagram
- `MU_SH0ES` : distance modulus
- `MU_SH0ES_ERR_DIAG` : uncertainty in the distance modulus

We then extract these cleaned columns as NumPy arrays to prepare for analysis and modeling.

In [3]:
```
z = df['zHD']
mu = df['MU_SH0ES']
mu_err = df['MU_SH0ES_ERR_DIAG']
```

## 📈 Plot the Hubble Diagram

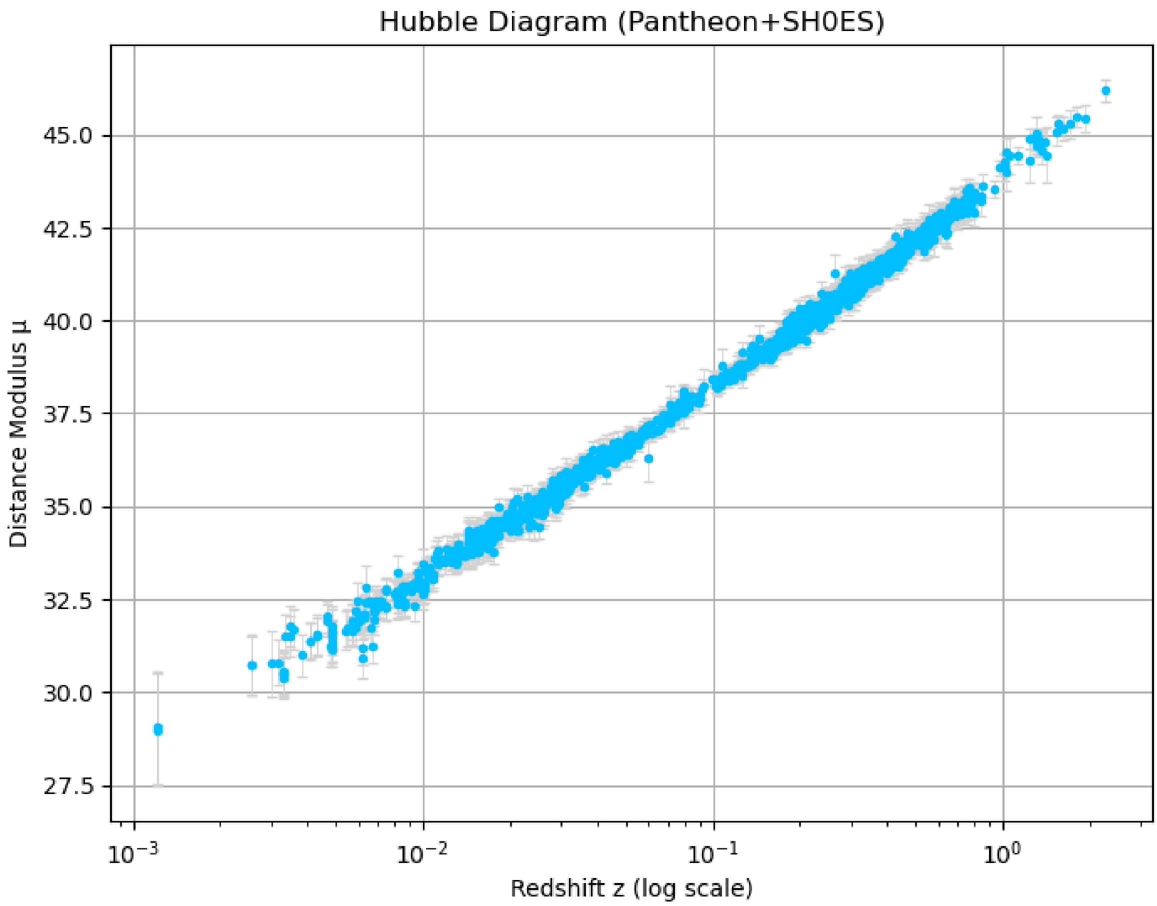
Let's visualize the relationship between redshift $z$ and distance modulus $\mu$, known as the Hubble diagram. This plot is a cornerstone of observational cosmology—it allows us to compare supernova observations with theoretical predictions based on different cosmological models.

We use a logarithmic scale on the redshift axis to clearly display both nearby and distant supernovae.

In [4]:
```
# Write a code to plot the distance modulus and the redshift (x-axis), label the

#Try using log scale in x-axis

plt.figure(figsize=(8,6))
plt.errorbar(z, mu, yerr=mu_err, fmt='.', color='deepskyblue', ecolor='lightgray
plt.xscale('log')
plt.xlabel("Redshift z (log scale)")
plt.ylabel("Distance Modulus μ")
plt.title("Hubble Diagram (Pantheon+SH0ES)")
plt.grid(True)
plt.show()
```




## 🧮 Define the Cosmological Model

We now define the theoretical framework based on the flat $\Lambda\mathrm{CDM}$ model (read about the model in wikipedia if needed). This involves:

- The dimensionless Hubble parameter:

$$E(z) = \sqrt{\Omega_m(1+z)^3 + (1-\Omega_m)}$$

- The distance modulus is:

$$\mu(z) = 5\log_{10}(d_L/\mathrm{Mpc}) + 25$$

- And the corresponding luminosity distance :

$$d_L(z) = (1+z) \cdot \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}$$

These equations allow us to compute the expected distance modulus from a given redshift $z$, Hubble constant $H_0$, and matter density parameter $\Omega_m$.

In [5]:

```
import numpy as np
from scipy.integrate import quad

# Constants
c = 3e5  # speed of light in km/s

# Cosmological model: flat Lambda-CDM
```

```python
def E(z, omega_m):
    return np.sqrt(omega_m * (1 + z)**3 + (1 - omega_m))

# Luminosity distance (in Mpc)
def luminosity_distance(z, H0, omega_m):
    integral, _ = quad(lambda z_: 1 / E(z_, omega_m), 0, z)
    return (1 + z) * (c / H0) * integral

# Distance modulus
def mu_theory(z, H0, omega_m):
    dL = luminosity_distance(z, H0, omega_m)
    return 5 * np.log10(dL) + 25
```

## 🔧 Fit the Model to Supernova Data

We now perform a non-linear least squares fit to the supernova data using our theoretical model for $\mu(z)$. This fitting procedure will estimate the best-fit values for the Hubble constant $H_0$ and matter density parameter $\Omega_m$, along with their associated uncertainties.

We'll use:

- `curve_fit` from `scipy.optimize` for the fitting.
- The observed distance modulus (\mu), redshift (z), and measurement errors.

The initial guess is:

- $H_0 = 70\,\mathrm{km/s/Mpc}$
- $\Omega_m = 0.3$

In [6]:
```python
# Initial guess: H0 = 70, Omega_m = 0.3
p0 = [70, 0.3]

# Wrapper model for curve_fit
def mu_fit(z, H0, omega_m):
    return np.array([mu_theory(zi, H0, omega_m) for zi in z])

# Fit the model
params, cov = curve_fit(mu_fit, z, mu, sigma=mu_err, p0=p0)

# Extract values and uncertainties
H0_fit, Omega_m_fit = params
H0_err, Omega_m_err = np.sqrt(np.diag(cov))

# Print the result
print(f"Fitted H0 = {H0_fit:.2f} ± {H0_err:.2f} km/s/Mpc")
print(f"Fitted Omega_m = {Omega_m_fit:.3f} ± {Omega_m_err:.3f}")
```

```
Fitted H0 = 73.02 ± 0.17 km/s/Mpc
Fitted Omega_m = 0.351 ± 0.012
```

## ⏳ Estimate the Age of the Universe

Now that we have the best-fit values of $H_0$ and $\Omega_m$, we can estimate the age of the universe. This is done by integrating the inverse of the Hubble parameter over redshift:

$$t_0 = \int_0^\infty \frac{1}{(1+z)H(z)}\,dz$$

We convert $H_0$ to SI units and express the result in gigayears (Gyr). This provides an independent check on our cosmological model by comparing the estimated age to values from other probes like Planck CMB measurements.

In [7]:

```
from scipy.integrate import quad
import numpy as np

def age_of_universe(H0, Omega_m):
    # Convert H0 from km/s/Mpc → 1/s
    H0_SI = H0 * 1000 / (3.086e22)  # 1/s
    Gyr = 3.154e16 # seconds in 1 Gyr

    # Define the integrand
    def integrand(z):
        Ez = np.sqrt(Omega_m * (1 + z)**3 + (1 - Omega_m))
        return 1 / ((1 + z) * Ez)

    # Integrate from 0 to 1000
    integral, _ = quad(integrand, 0, 1000)
    print("Integral:", integral)  # DEBUG

    # Calculate age
    t0_seconds = integral / H0_SI
    t0_gyr = t0_seconds / Gyr
    return t0_gyr

t0 = age_of_universe(H0_fit, Omega_m_fit)
print(f"Estimated age of Universe: {t0:.2f} Gyr")
```

```
Integral: 0.9222903108775391
Estimated age of Universe: 12.36 Gyr
```

# 📊 Analyze Residuals

To evaluate how well our cosmological model fits the data, we compute the residuals:

$$\text{Residual} = \mu_{\text{obs}} - \mu_{\text{model}}$$

Plotting these residuals against redshift helps identify any systematic trends, biases, or outliers. A good model fit should show residuals scattered randomly around zero without any significant structure.

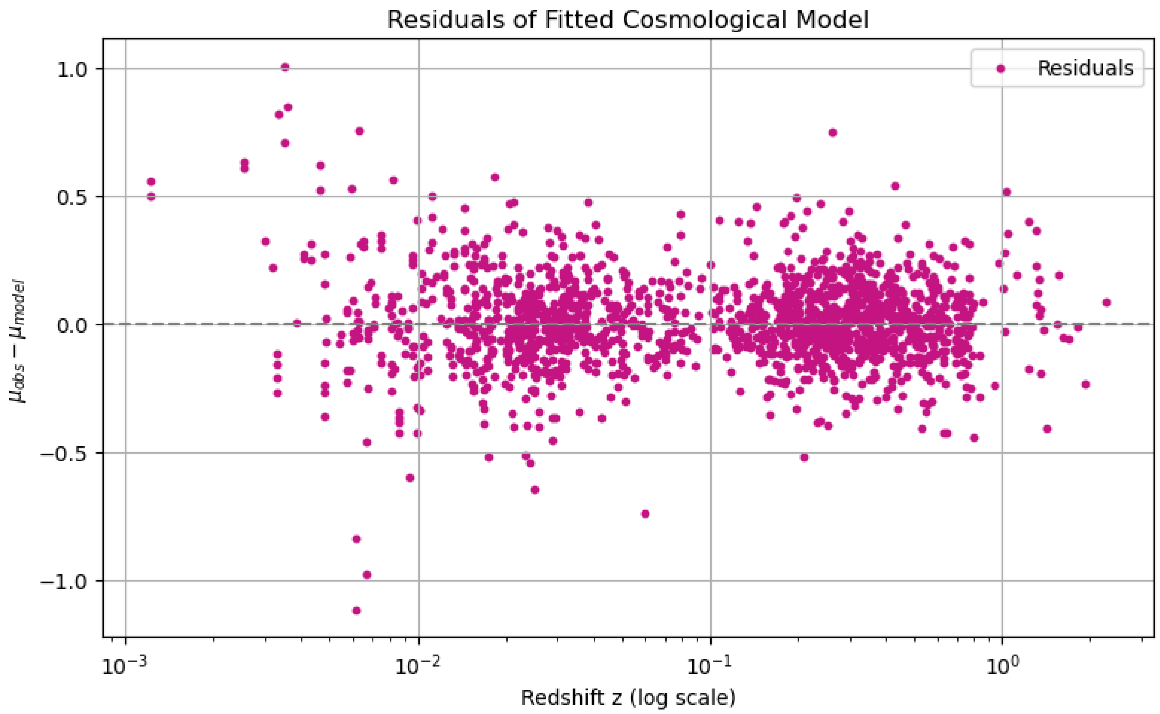
In [8]:

```
# Calculate model predictions
mu_model = np.array([mu_theory(zi, H0_fit, Omega_m_fit) for zi in z])

# Calculate residuals
residuals = mu - mu_model

# Plot
plt.figure(figsize=(8, 5))
```

```
plt.scatter(z, residuals, s=10, color='mediumvioletred', label='Residuals')
plt.axhline(0, color='gray', linestyle='--')
plt.xscale('log')
plt.xlabel("Redshift z (log scale)")
plt.ylabel(r"$\mu_{obs} - \mu_{model}$")
plt.title("Residuals of Fitted Cosmological Model")
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```




## 🔧 Fit with Fixed Matter Density

To reduce parameter degeneracy, let's fix $\Omega_m = 0.3$ and fit only for the Hubble constant $H_0$.

In [9]:
```
def mu_fixed_Om(z_array, H0):
    return np.array([mu_theory(zi, H0, 0.3) for zi in z_array])

# Initial guess for H0
p0_fixed = [70]

# Fit only H0 while fixing Omega_m = 0.3
params_fixed, cov_fixed = curve_fit(mu_fixed_Om, z, mu, sigma=mu_err, p0=p0_fixe

# Extract result
H0_fixed = params_fixed[0]
H0_fixed_err = np.sqrt(cov_fixed[0, 0])

print(f"Fitted H0 with Ωm=0.3: {H0_fixed:.2f} ± {H0_fixed_err:.2f} km/s/Mpc")
```

```
Fitted H0 with Ωm=0.3: 73.58 ± 0.11 km/s/Mpc
```

## 🔍 Compare Low-z and High-z Subsamples

Finally, we examine whether the inferred value of $H_0$ changes with redshift by splitting the dataset into:

- **Low-z** supernovae ($z < 0.1$)
- **High-z** supernovae ($z \geq 0.1$)

We then fit each subset separately (keeping $\Omega_m = 0.3$) to explore any potential tension or trend with redshift.

In [10]:
```
# Low-z: z < 0.1
mask_low = z < 0.1
z_low = z[mask_low]
mu_low = mu[mask_low]
mu_err_low = mu_err[mask_low]

# High-z: z >= 0.1
mask_high = z >= 0.1
z_high = z[mask_high]
mu_high = mu[mask_high]
mu_err_high = mu_err[mask_high]

# LOW-z fit
params_low, cov_low = curve_fit(mu_fixed_Om, z_low, mu_low, sigma=mu_err_low, p0
H0_low = params_low[0]
H0_low_err = np.sqrt(cov_low[0][0])

# HIGH-z fit
params_high, cov_high = curve_fit(mu_fixed_Om, z_high, mu_high, sigma=mu_err_hig
H0_high = params_high[0]
H0_high_err = np.sqrt(cov_high[0][0])

print(f"Low-z (z < 0.1): H0 = {H0_low:.2f} ± {H0_low_err:.2f} km/s/Mpc")
print(f"High-z (z ≥ 0.1): H0 = {H0_high:.2f} ± {H0_high_err:.2f} km/s/Mpc")
```

```
Low-z (z < 0.1): H0 = 73.06 ± 0.19 km/s/Mpc
High-z (z ≥ 0.1): H0 = 73.90 ± 0.14 km/s/Mpc
```

You can check your results and potential reasons for different values from accepted constant using this paper by authors of the Pantheon+ dataset

You can find more about the dataset in the paper too