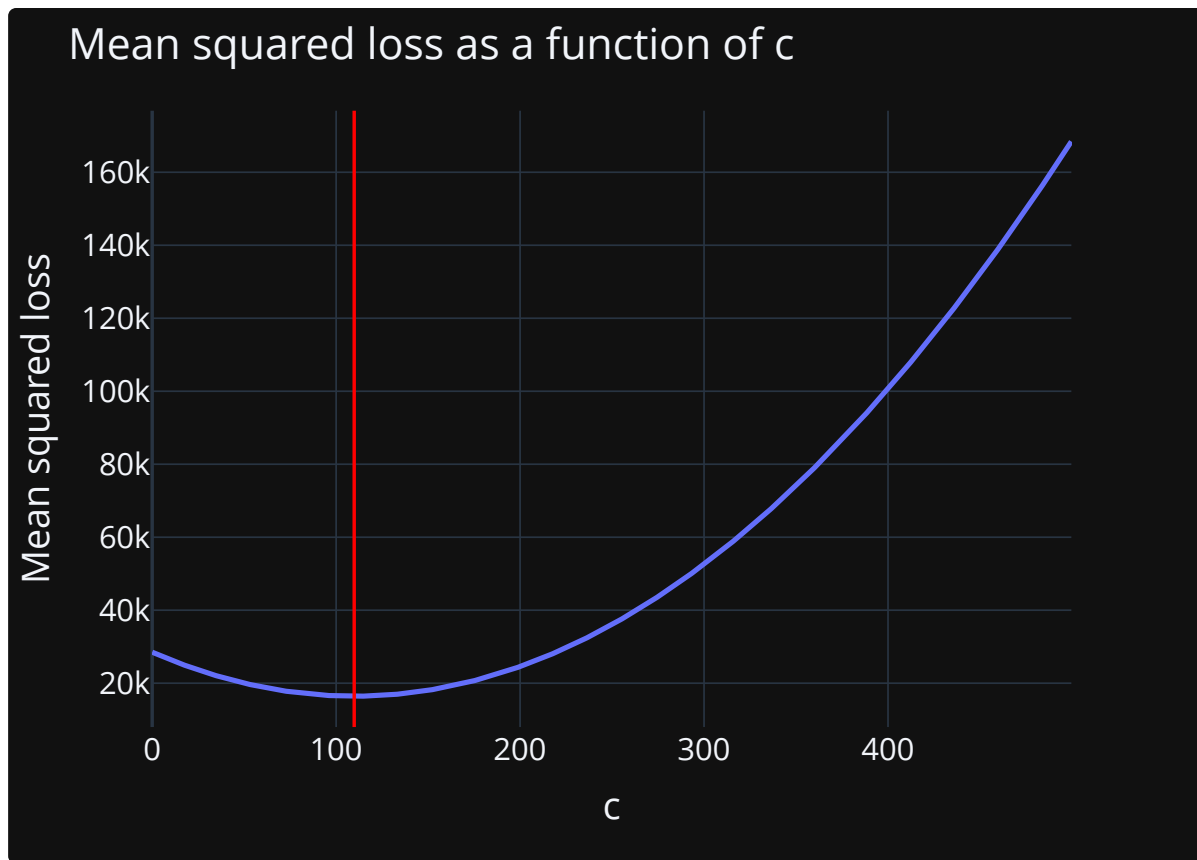


Applied Data Analytics

Pandas basics

Simulating data

Hans-Martin von Gaudecker and Aapo Stenhammar



Simulating data

- Called **Monte Carlo** studies because of randomness
- Will see mathematical basis only in statistics course
- Just work with intuition & recipe here

Recipe for simulating a series

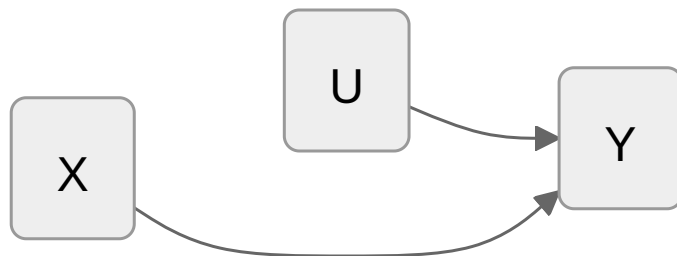
```
rng = np.random.default_rng(seed=243345)

samples = pd.Series(
    rng.normal(loc=10, scale=5, size=10),
    name="u"
)

print(samples)
```

	u
0	8.32006
1	13.5407
2	18.185
3	2.93449
4	9.22757
5	17.8104
6	12.8181
7	10.3423
8	9.85138
9	12.4335

OLS, happy path



OLS, happy path

$$X_i \sim \text{Normal}(10, 5)$$

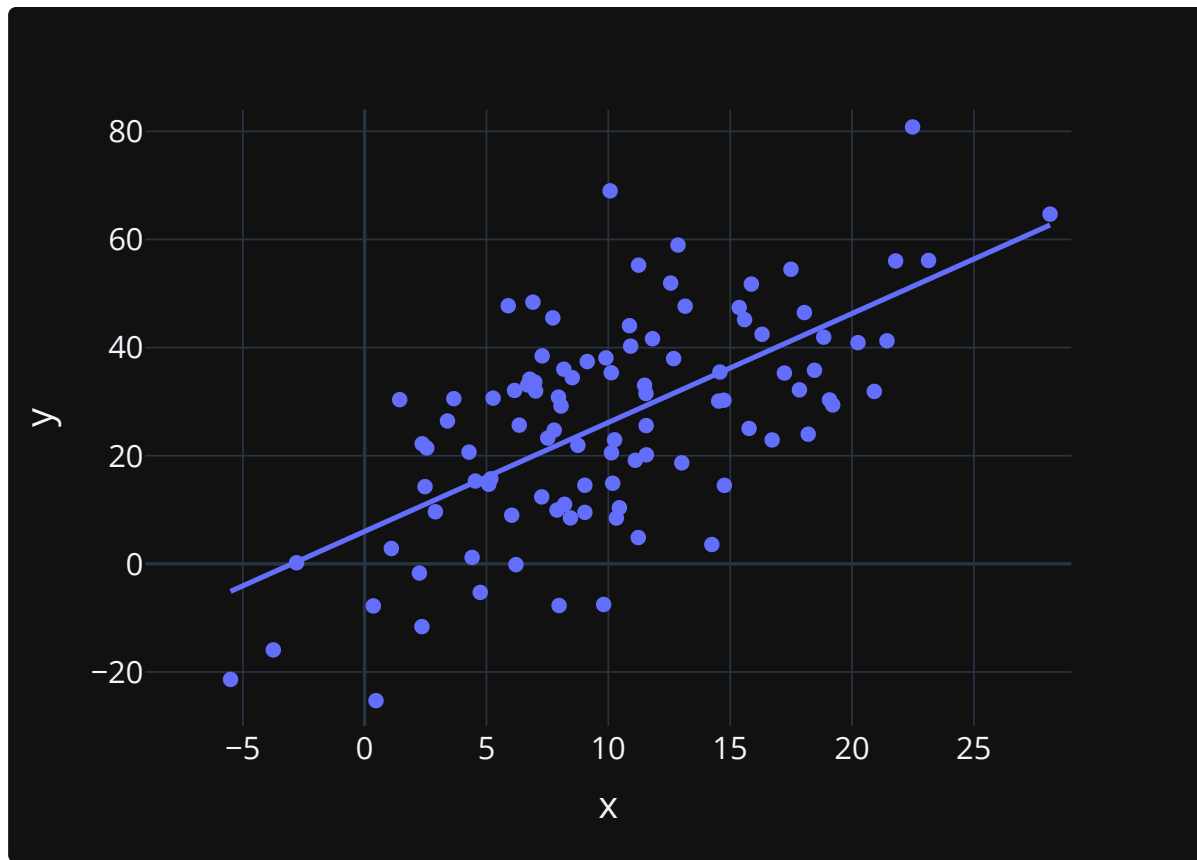
$$U_i \sim \text{Normal}(0, 15)$$

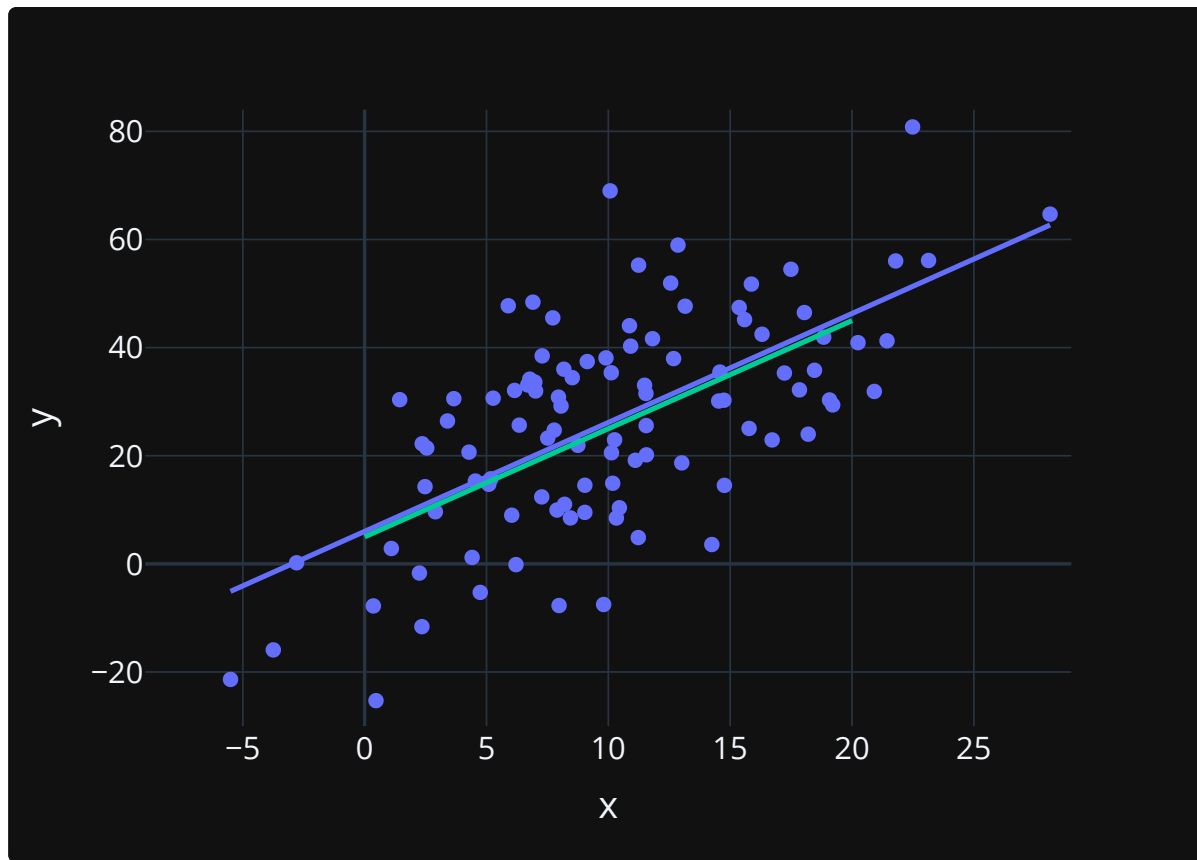
$$Y_i = 5 + 2 \cdot X_i + U_i$$

```
size = 100
```

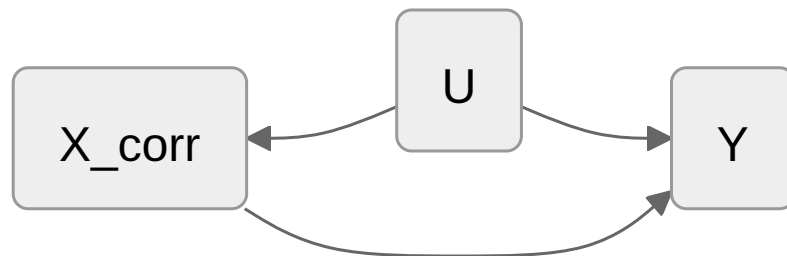
```
data = pd.DataFrame({  
    "x": rng.normal(loc=10, scale=5, size=size),  
    "u": rng.normal(loc=0, scale=15, size=size),  
})
```

```
data["y"] = 5 + 2 * data["x"] + data["u"]
```





OLS, with confounder



OLS, with confounder

```
data_corr = data.copy()
```

$$X_i^{\text{corr}} = X_i + U_i/3$$

```
data_corr["x"] = data_corr["x"] - data_corr["u"] / 3
```

$$Y_i^{\text{corr}} = 5 + 2 \cdot X_i^{\text{corr}} + U_i$$

```
data_corr["y"] = 5 + 2 * data_corr["x"] + data_corr["u"]
```

