

Model card

Linear and logistic regression in statsmodels, reading the output, the four checks.

```
import statsmodels.formula.api as smf

lin = smf.ols("los ~ age + emergency + ward", data=df).fit(); lin.summary()

log = smf.logit("died ~ age + emergency + ward", data=df).fit()

np.exp(log.params); np.exp(log.conf_int())    # odds ratios and intervals
```

Check	What it protects against
Residuals against predicted values	A curved relationship modelled as a line
Spread of residuals	Fanning out, common with length of stay
Influence, refit without extreme cases	One patient driving the result
Independence, patients within wards or clinicians	Overconfident intervals, consider clustered or mixed models

- About ten outcome events per predictor. Count the events before planning the model.
- Report the crude estimate beside the adjusted one and name the reference category for every categorical variable.
- An observational model estimates association after adjustment. It does not prove cause.

Based on the ERAS Africa Advanced Data Analysis Using Python course, 2026, WHO surveillance guidance, the EPHI public health emergency management guidelines, the STROBE statement and the pandas, SciPy and statsmodels documentation. Free to copy and adapt with attribution.