

Quantitative Reasoning – LAB 413

Inference for Regression, Multivariate Regression & Interaction Terms

Week 11 Recap ▪ Lab 10

Instructor: **Subin Na**

November 12, 2025

Today's Plan

1. Quick Recap

- Inference for Proportions (review)
- Inference for Regression
- OLS Assumptions
- Multiple Regression Concept

2. RStudio Practice

01 Inference for Proportions (1)

- We use sample proportions $\hat{p}$ to estimate the population proportion p .
- Because samples vary, $\hat{p}$ has its own **sampling distribution** → roughly Normal if:
 - Random sample
 - Expected successes ≥ 10 , failures ≥ 10

Confidence Interval (CI)

A range of plausible population proportions:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

01 Inference for Proportions (2)

Example

In a sports study, **73 out of 150** athletes reported sleeping more than 8 hours the previous night. Compute and interpret a **90% confidence interval** for this proportion.

- Point estimate: $\hat{p} = \frac{73}{150} \approx 0.487$.
- Critical value for a 90% CI: $z^* = 1.645$.

Inference for Proportions: Solution

Plug in and solve:

$$\begin{aligned}\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} &= 0.487 \pm 1.645 \sqrt{\frac{(0.487)(1 - 0.487)}{150}} \\ &= 0.487 \pm 0.067\end{aligned}$$

90% CI is (0.420, 0.554)

02 Inference for Regression

- Regression estimates $(\hat{\beta}_0, \hat{\beta}_1)$ are statistics computed from samples.
- Each has a **sampling distribution** $\rightarrow$ uncertainty measured by the standard error (SE).

Key Questions

- Is the relationship **real** or due to chance?
 - How **precise** is our slope estimate?
-
- 95% CI $\approx \hat{\beta} \pm 1.96 \times \text{SE}$.
 - Reject H_0 if $p < 0.05$ (β significantly different from 0).

03 OLS Assumptions

- **Linearity** — relationship between X and Y is linear.
- **Random sampling** — observations are independent and random.
- **Exogeneity** — errors uncorrelated with predictors.
- **Homoskedasticity** — constant error variance.
- **Variation in X** — predictors are not constant.

(Violating these assumptions $\Rightarrow$ biased or inefficient estimates.)

04 Multiple Regression Concept

Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \varepsilon$$

- Adds more information and **controls for confounding**.
- Each coefficient is interpreted *holding other variables constant*.
- Helps separate true effects from spurious associations.
- Addresses **omitted variable bias** if relevant controls are included.

RStudio Practice

Proportion Tests & Inference for Regression

See `Lab10.Rmd` / `Lab10_1112.R`