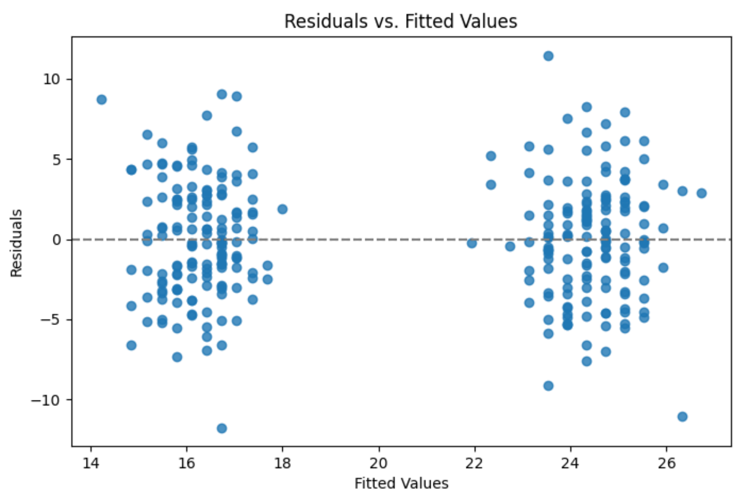
[STATS 413] HW 6

Huy Le

February 2026

1 Problem 1

(a) Residuals versus fitted values plot:



(b) From the residuals-versus-fitted plot, the residuals appear to be roughly centered around zero with no strong systematic trend, which supports the linearity assumption. The vertical spread of residuals looks approximately constant across the range of fitted values, suggesting that homoskedasticity is at least reasonably plausible. There are a few points that could be moderately unusual (larger positive/negative residuals). Overall, the stronger linear model assumptions appear reasonable based on this diagnostic plot.

(c) A 95% interval for the expected income (in thousands of dollars) for Millenial residents with 8 years of education is

$$(14.59,\ 16.37),$$

with point estimate $\hat{\mu} = 15.48$.

2 Problem 2

(a) $\hat{\mu}_{y|c} = c^\top \hat{\beta}$.

(b) $\hat{\mu}_{y|c} - \hat{\mu}_{y|d} = (c-d)^\top \hat{\beta}$.

(c) Since $\mathbb{E}(\hat{\beta}) = \beta$, we have

$$\mathbb{E}\big[\hat{\mu}_{y|c} - \hat{\mu}_{y|d}\big] = (c-d)^\top \mathbb{E}(\hat{\beta}) = (c-d)^\top \beta.$$

(d) Note that $\mathrm{Var}(\hat{\beta}) = \sigma_\varepsilon^2 (X^\top X)^{-1}$, we have

$$\mathrm{Var}(\hat{\mu}_{y|c} - \hat{\mu}_{y|d}) = \mathrm{Var}\big((c-d)^\top \hat{\beta}\big) = (c-d)^\top \mathrm{Var}(\hat{\beta})(c-d) = \sigma_\varepsilon^2 (c-d)^\top (X^\top X)^{-1}(c-d).$$

(e)

$$E\{y_c^* - y_d^* - (\hat{\mu}_{y|c} - \hat{\mu}_{y|d})\} = E\{(c-d)^\top \beta + (\varepsilon_c^* - \varepsilon_d^*) - (c-d)^\top \hat{\beta}\} = (c-d)^\top (\beta - \mathbb{E}\hat{\beta}) + E(\varepsilon_c^* - \varepsilon_d^*) = 0.$$

(f) Since $(\varepsilon_c^* - \varepsilon_d^*)$ is independent of $\hat{\beta}$ and has variance $2\sigma_\varepsilon^2$,

$$\mathrm{Var}\{y_c^* - y_d^* - (\hat{\mu}_{y|c} - \hat{\mu}_{y|d})\} = \mathrm{Var}(\varepsilon_c^* - \varepsilon_d^*) + \mathrm{Var}\big((c-d)^\top \hat{\beta}\big) = 2\sigma_\varepsilon^2 + \sigma_\varepsilon^2 (c-d)^\top (X^\top X)^{-1}(c-d).$$

(g) A 95% confidence interval for the difference in incomes between person c and person d is (-18.6773, 1.7944).

(h) A 95% confidence interval for the difference in incomes between two people with the same values of Millenial and Education is (-10.1738, 10.1738).