
Solution 1

(a) The column '`t value`"gives the t-statistics to test the null hypotheses $\beta_r = 0$, i.e.,

$$T_r = \frac{\widehat{\beta}_r}{\sqrt{S^2 v_{rr}}} = \frac{\widehat{\beta}_r}{\widehat{\mathrm{SE}}(\widehat{\beta}_r)},$$

where v_{rr} is the rth diagonal element of the matrix $(X^{\mathrm{T}}X)^{-1}$. When in fact $\beta_r = 0$, $T_r \sim t_{n-p}$, and the null hypothesis is rejected for large $|T_r|$.

The column '`Pr(>|t|)`' gives the p-values for two-sided t tests of these hypotheses. For an observed statistic $t_{r,\mathrm{obs}}$, the p-value is

$$p_r = \mathrm{P}(|T_r| > |t_{r,\mathrm{obs}}|) = 2\{1 - F_{n-p}(|t_{r,\mathrm{obs}}|)\} = 2F_{n-p}(-|t_{r,\mathrm{obs}}|),$$

where F_{n-p} is the Student t CDF with $n-p$ degrees of freedom.

If $p_r < 0.05$, for example, we reject rth hypothesis at the 5% significance level. For this example, with 5% significance level, we can reject $\beta_r = 0$ for $r = 0, 1, 2$, but not for $r = 3$.

(b) Computing the variance formula simply uses the properties of covariance and correlation.

The test statistic is

$$T = \frac{c^{\mathrm{T}}\widehat{\beta} - c^{\mathrm{T}}\beta}{\sqrt{S^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}} = \frac{c^{\mathrm{T}}\widehat{\beta}}{\sqrt{S^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c}}$$

for $c = [0, 0, 1, -1]^{\mathrm{T}}$ and under the hypothesis that $c^{\mathrm{T}}\beta = 0$. Now

$$\begin{aligned} s^2 c^{\mathrm{T}}(X^{\mathrm{T}}X)^{-1}c &= \left\{\mathrm{SE}\left(\widehat{\beta}_2\right)\right\}^2 + \left\{\mathrm{SE}\left(\widehat{\beta}_3\right)\right\}^2 - 2\,\mathrm{corr}\left(\widehat{\beta}_2, \widehat{\beta}_3\right)\,\mathrm{SE}\left(\widehat{\beta}_2\right)\,\mathrm{SE}\left(\widehat{\beta}_3\right) \\ &= 0.04423^2 + 0.18471^2 - 2 \cdot (-0.08911) \cdot 0.04423 \cdot 0.18471 = 0.03753, \end{aligned}$$

so

$$T = \frac{0.65691 - 0.25002}{\sqrt{0.03753}} = 2.10033,$$

and the corresponding p-value is

$$p = 2F_{13-4}(-2.10033) = 0.06508.$$

Thus we do not reject the null hypothesis at the 5% level. This seems a bit surprising at first sight, because the estimates are rather different and almost uncorrelated, but on the other hand the standard error for $\widehat{\beta}_3$ is rather large, even if that for $\widehat{\beta}_2$ is much smaller.

Solution 2

(a) (i) $X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$, $\beta = (\beta_0, \alpha_1, \alpha_2)^{\mathrm{T}}$.

(ii) Clearly X has rank 2 and therefore $X^{\mathrm{T}}X$ is not invertible. Thus all three parameters cannot be estimated separately, i.e., the model is not identifiable.

(iii) Here

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Suppressing the column for α_1 is equivalent to setting $\alpha_1 = 0$. β_0 is the mean of the observations for which $a_j = 1$ and α_2 is the difference between the means of the groups with $a_j = 1$ and $a_j = 2$.

(iv) For the model $y \sim a$,

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

For the model $y \sim a + b$,

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

For the model $y \sim x + a - 1$,

$$X = \begin{pmatrix} 152 & 1 & 0 \\ 93 & 1 & 0 \\ 127 & 1 & 0 \\ 109 & 0 & 1 \\ 141 & 0 & 1 \\ 136 & 0 & 1 \end{pmatrix}.$$

For the model $y \sim b + x - 1$,

$$X = \begin{pmatrix} 1 & 0 & 0 & 152 \\ 0 & 1 & 0 & 93 \\ 0 & 0 & 1 & 127 \\ 1 & 0 & 0 & 109 \\ 0 & 1 & 0 & 141 \\ 0 & 0 & 1 & 136 \end{pmatrix}.$$

(b) For the model $y \sim a : x$,

$$X = \begin{pmatrix} 1 & 152 & 0 \\ 1 & 93 & 0 \\ 1 & 127 & 0 \\ 1 & 0 & 109 \\ 1 & 0 & 141 \\ 1 & 0 & 136 \end{pmatrix};$$

the columns are linearly independent.

For the model $y \sim a : b$,

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix};$$

the columns are not linearly independent.

For the model $y \sim a + b : x$,

$$X = \begin{pmatrix} 1 & 0 & 152 & 0 & 0 \\ 1 & 0 & 0 & 93 & 0 \\ 1 & 0 & 0 & 0 & 127 \\ 1 & 1 & 109 & 0 & 0 \\ 1 & 1 & 0 & 141 & 0 \\ 1 & 1 & 0 & 0 & 136 \end{pmatrix};$$

the columns are linearly independent.

For the model $y \sim a + a : b : x$,

$$X = \begin{pmatrix} 1 & 0 & 152 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 93 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 127 & 0 \\ 1 & 1 & 0 & 109 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 141 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 136 \end{pmatrix};$$

the columns are not linearly independent.

Solution 3

(a) If the Gaussian model is correct, then the residuals e and the fitted values $\widehat{y}$ are independent and the standardized residuals r_j have a standard normal distribution; hence we expect to see 95% of the $\|r_j| < 2$, and $r \perp\!\!\!\perp y$.

- Plot A: the model seems reasonable.
- Plot B: there is an outlier with $r_j < -2$. We should check whether there is something special about this observation, and see whether omitting it changes the fit and interpretation appreciably.
- Plot C: the fitted values and standardized residuals seem dependent. The addition of a quadratic term in one of the covariates might solve this.
- Plot D: the variance of the residual grows with the fitted mean. Weighted least squares estimation might be preferable.

(b) If the data distribution has a lower tail (left side of the distribution) heavier than the normal law, the empirical quantiles on the lower left part of the quantile-quantile plot will be under the diagonal $y = x$. Indeed in this case, $G(x) \gg \Phi(x)$ when $x \to -\infty$, where G is the empirical distribution function and ϕ the Gaussian distribution function. For small $\alpha > 0$, if

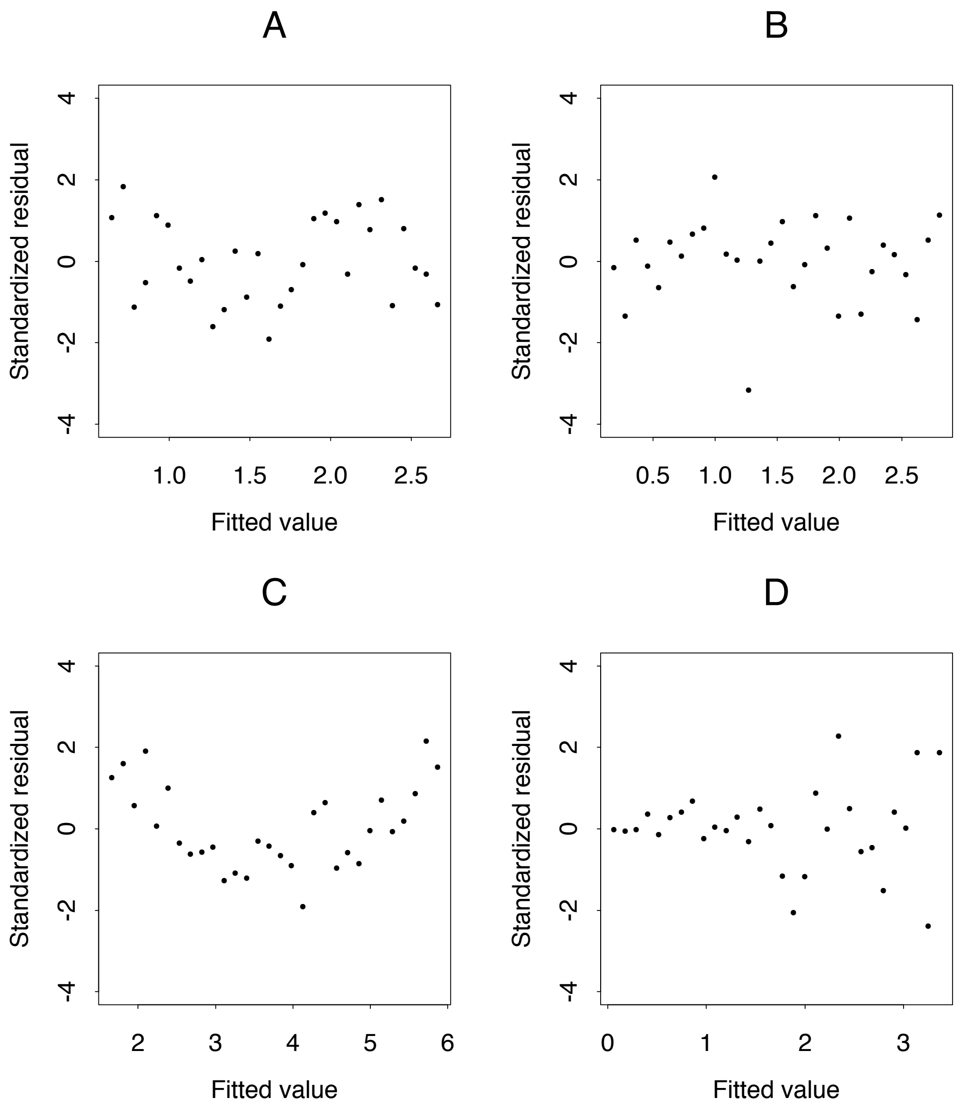

Figure 1: Standardized residuals for four Gaussian linear models.

$x = F^{-1}(\alpha)$, then $G(x) \gg \alpha$, and so $G^{-1}(\alpha) < x$, which means that the points will be under the line $y = x$. On the contrary, if the lower tail is lighter than the normal distribution, then empirical quantiles will be above the diagonal. Similar considerations give the behaviour for the upper (right) tail.

- Plot A: heavy lower tail and light upper tail, showing negative skewness.
- Plot B: tails are lighter than Gaussian.
- Plot C: tails are heavier than Gaussian
- Plot D: light lower tail and heavy upper tail, showing positive skewness.

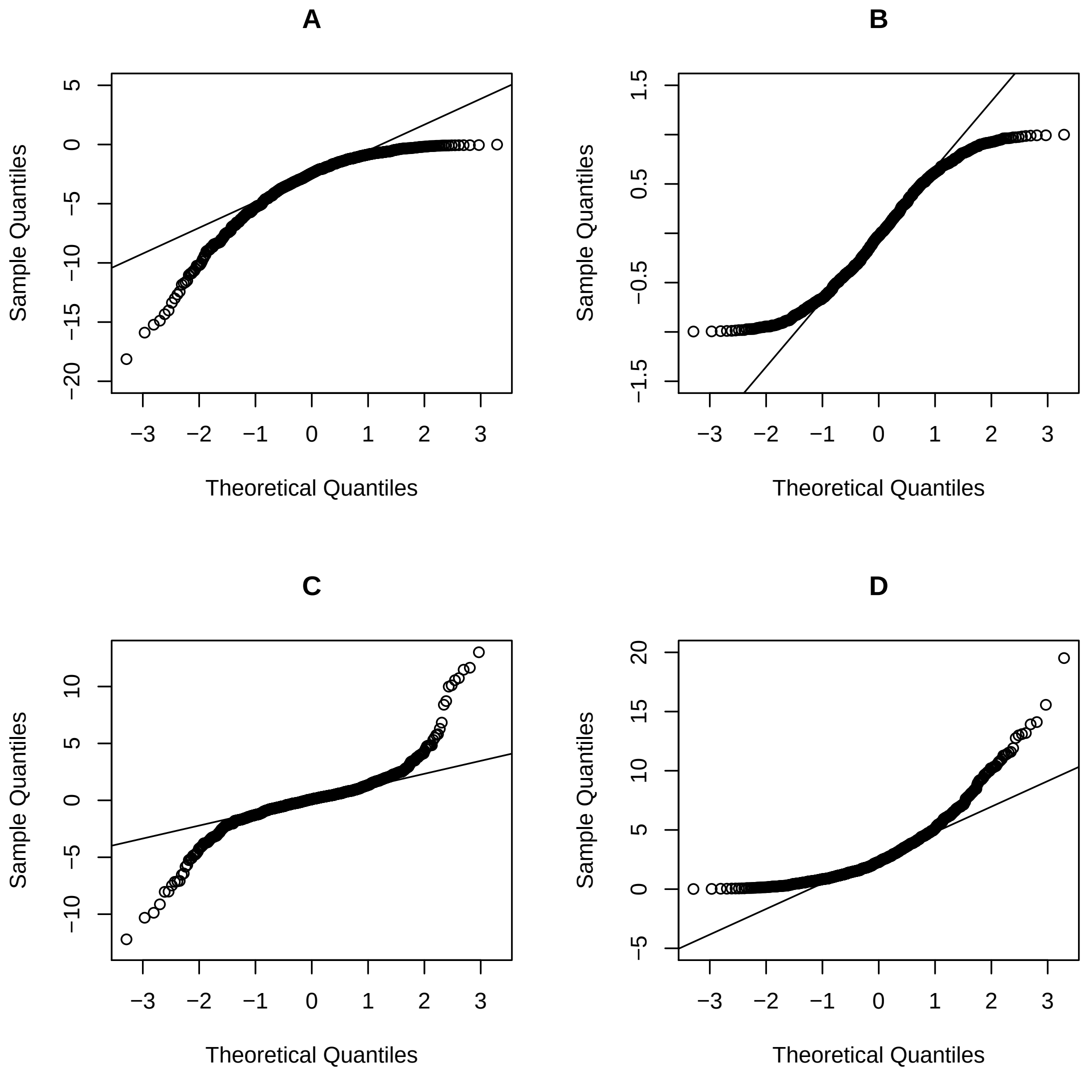

Figure 2: Four Gaussian Q-Q plots.