Lecture 14. Applications of the distribution theory

Inference for $\boldsymbol{\beta}$

We know that $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(X^T X)^{-1})$, and so

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2(X^T X)^{-1}_{jj}).$$

The *standard error* of $\hat{\boldsymbol{\beta}}_j$ is

$$\text{s.e.}(\hat{\beta}_j) = \sqrt{\tilde{\sigma}^2(X^T X)^{-1}_{jj}},$$

where $\tilde{\sigma}^2 = \text{RSS}/(n-p)$, as in Theorem 13.3.

Then

$$\frac{\hat{\beta}_j - \beta_j}{\text{s.e.}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\tilde{\sigma}^2(X^T X)^{-1}_{jj}}} = \frac{(\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2(X^T X)^{-1}_{jj}}}{\sqrt{\text{RSS}/((n-p)\sigma^2)}}.$$

The numerator is a standard normal $N(0,1)$, the denominator is an independent $\sqrt{\chi^2_{n-p}/(n-p)}$, and so $\frac{\hat{\beta}_j - \beta_j}{\text{s.e.}(\hat{\beta}_j)} \sim t_{n-p}$.

So a $100(1-\alpha)\%$ CI for β_j has endpoints $\hat{\beta}_j \pm \text{s.e.}(\hat{\beta}_j)\ t_{n-p}(\frac{\alpha}{2})$.

To test $H_0 : \beta_j = 0$, use the fact that, under H_0, $\frac{\hat{\beta}_j}{\text{s.e.}(\hat{\beta}_j)} \sim \mathrm{t}_{n-p}$.

Simple linear regression

We assume that

$$Y_i = a' + b(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where $\bar{x} = \sum x_i/n$, and $\varepsilon_i, i = 1, ..., n$ are iid $\mathrm{N}(0, \sigma^2)$.

Then from Lecture 12 and Theorem 13.3 we have that

$$\hat{a}' = \overline{Y} \sim \mathrm{N}\left(a', \frac{\sigma^2}{n}\right), \qquad \hat{b} = \frac{S_{xY}}{S_{xx}} \sim \mathrm{N}\left(b, \frac{\sigma^2}{S_{xx}}\right),$$

$$\hat{Y}_i = \hat{a}' + \hat{b}(x_i - \bar{x}), \qquad \mathrm{RSS} = \sum_i (Y_i - \hat{Y}_i)^2 \sim \sigma^2 \chi^2_{n-2},$$

and $(\hat{a}', \hat{b})$ and $\hat{\sigma}^2 = \mathrm{RSS}/n$ are independent.

Example 12.1 continued

- We have seen that $\tilde{\sigma}^2 = \frac{\text{RSS}}{n-p} = \frac{67968}{(24-2)} = 3089 = 55.6^2$.
- So the standard error of $\hat{b}$ is

$$\text{s.e.}(\hat{b}) = \sqrt{\tilde{\sigma}^2 (X^T X)^{-1}_{22}}, = \sqrt{\frac{3089}{S_{xx}}} = \frac{55.6}{28.0} = 1.99.$$

- So a 95% interval for b has endpoints $\hat{b} \pm \text{s.e.}(\hat{b}) \times t_{n-p}(0.025) = -12.9 \pm 1.99 * t_{22}(0.025) = (-17.0, -8.8)$, where $t_{22}(0.025) = 2.07$.
- This does not contain 0. Hence if carry out a size 0.05 test of $H_0 : b = 0$ *vs* $H_1 : b \neq 0$, the test statistic would be $\frac{\hat{b}}{\text{s.e.}(\hat{b})} = \frac{-12.9}{1.99} = -6.48$, and we would reject H_0 since this is less than $-t_{22}(0.025) = -2.07$.
-

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  826.500     11.346  72.846  < 2e-16 ***
oxy.s        -12.869      1.986  -6.479 1.62e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 55.58 on 22 degrees of freedom
```

Expected response at $\mathbf{x}^*$

- Let $\mathbf{x}^*$ be a new vector of values for the explanatory variables
- The expected response at $\mathbf{x}^*$ is $\mathbb{E}(Y|\mathbf{x}^*) = \mathbf{x}^{*T}\boldsymbol{\beta}$.
- We estimate this by $\mathbf{x}^{*T}\hat{\boldsymbol{\beta}}$.
- By Theorem 13.3 and Proposition 11.1(i),

$$\mathbf{x}^{*T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \mathrm{N}(0, \sigma^2\, \mathbf{x}^{*T}(X^T X)^{-1}\mathbf{x}^*).$$

- Let $\tau^2 = \mathbf{x}^{*T}(X^T X)^{-1}\mathbf{x}^*$.
- Then

$$\frac{\mathbf{x}^{*T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\tilde{\sigma}\tau} \sim t_{n-p}.$$

- A $100(1-\alpha)\%$ confidence interval for the expected response $\mathbf{x}^{*T}\boldsymbol{\beta}$ has endpoints

$$\mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \pm \tilde{\sigma}\tau\, t_{n-p}(\tfrac{\alpha}{2}).$$

Example 12.1 continued

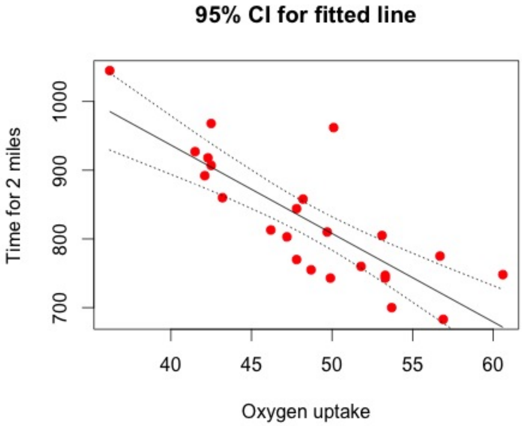
- Suppose we wish to estimate the time to run 2 miles for a man with an oxygen take-up measurement of 50.
- Here $\mathbf{x}^{*T} = (1, (50 - \bar{x}))$, where $\bar{x} = 48.6$.
- The estimated expected response at $\mathbf{x}^{*T}$ is

$$\mathbf{x}^{*T}\hat{\beta} = \hat{a}' + (50 - 48.6) \times \hat{b} = 826.5 - 1.4 \times 12.9 = 808.5.$$

- We find $\tau^2 = \mathbf{x}^{*T}(X^TX)^{-1}\mathbf{x}^* = \frac{1}{n} + \frac{\mathbf{x}^{*2}}{S_{xx}} = \frac{1}{24} + \frac{1.4^2}{783.5} = 0.044 = 0.21^2$.
- So a 95% CI for $\mathbb{E}(Y|\mathbf{x}^* = 50 - \bar{x})$ is

$$\mathbf{x}^{*T}\hat{\beta} \pm \tilde{\sigma}\tau t_{n-p}(\tfrac{\alpha}{2}) = 808.5 \pm 55.6 \times 0.21 \times 2.07 = (783.6, 832.2).$$

```
oxy.s = oxy - mean(oxy)
fit=lm(time~ oxy.s  )
pred=predict.lm(fit,  interval="confidence")
plot(oxy, time,col="red", pch=19, xlab="Oxygen uptake",ylab="Time for 2 miles", main
lines(oxy, pred[, "fit"])
lines(oxy, pred[, "lwr"], lty = "dotted")
lines(oxy, pred[, "upr"], lty = "dotted")
```



Predicted response at $x*$

- The response at $\mathbf{x}^*$ is $Y^* = \mathbf{x}^* \boldsymbol{\beta} + \varepsilon^*$, where $\varepsilon^* \sim \mathrm{N}(0, \sigma^2)$, and Y^* is independent of $Y_1, .., Y_n$.
- We predict $\hat{Y}^*$ by $\mathbf{x}^{*T} \hat{\boldsymbol{\beta}}$.
- A $100(1-\alpha)\%$ *prediction interval* for Y^* is an interval $I(\mathbf{Y})$ such that $\mathbb{P}(Y^* \in I(\mathbf{Y})) = 1 - \alpha$, where the probability is over the joint distribution of $(Y^*, Y_1, ..., Y_n)$.
- Observe that $\hat{Y}^* - Y^* = \mathbf{x}^{*T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \varepsilon^*$.
- So $\mathbb{E}(\hat{Y}^* - Y^*) = \mathbf{x}^{*T}(\boldsymbol{\beta} - \boldsymbol{\beta}) = 0$.
- And

$$\begin{aligned} \mathrm{var}(\hat{Y}^* - Y^*) &= \mathrm{var}(\mathbf{x}^{*T}(\hat{\boldsymbol{\beta}})) + \mathrm{var}(\varepsilon^*) \\ &= \sigma^2 \mathbf{x}^{*T}(X^T X)^{-1} \mathbf{x}^* + \sigma^2 \\ &= \sigma^2(\tau^2 + 1) \end{aligned}$$

- So

$$\hat{Y}^* - Y^* \sim \mathrm{N}(0, \sigma^2(\tau^2 + 1)).$$

- We therefore find that

$$\frac{\hat{Y}^* - Y^*}{\tilde{\sigma}\sqrt{(\tau^2+1)}} \sim t_{n-p}.$$

- So the interval with endpoints

$$\mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \pm \tilde{\sigma}\sqrt{(\tau^2+1)}\; t_{n-p}(\tfrac{\alpha}{2}).$$

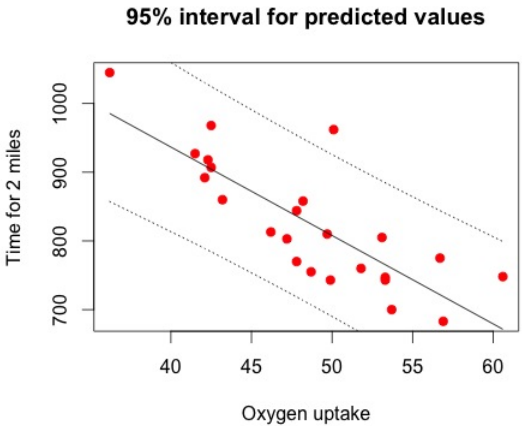
  is a 95% prediction interval for Y^*.

Example 12.1 continued

A 95% prediction interval for Y^* at $\mathbf{x}^{*T} = (1, (50-\bar{x}))$ is

$$\mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \pm \tilde{\sigma}\sqrt{(\tau^2+1)}\; t_{n-p}(\tfrac{\alpha}{2}) = 808.5 \pm 55.6 \times 1.02 \times 2.07 = (691.1, 925.8).$$

```
pred=predict.lm(fit,  interval="prediction")
```



Note wide prediction intervals for individual points, with the width of the interval dominated by the residual error term $\tilde{\sigma}$ rather than the uncertainty about the fitted line.

Example 13.1 continued. One-way analysis of variance

- Suppose we wish to estimate the expected resistivity of a new wafer in the first instrument.
- Here $\mathbf{x}^{*T} = (1, 0, .., 0)$.
- The estimated expected response at $\mathbf{x}^{*T}$ is

$$\mathbf{x}^{*T}\hat{\boldsymbol{\mu}} = \hat{\mu}_1 = \overline{Y}_{1.} = 124.3$$

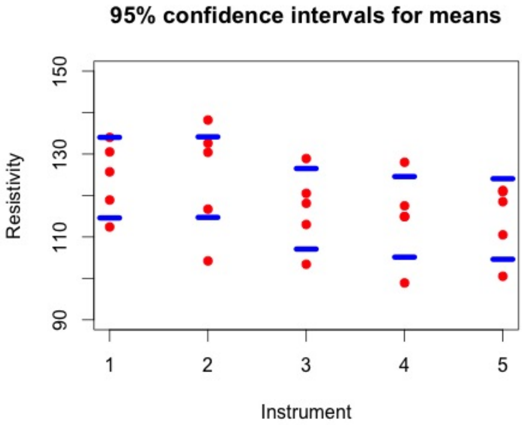
- We find $\tau^2 = \mathbf{x}^{*T}(X^T X)^{-1}\mathbf{x}^* = \frac{1}{5}$.
- So a 95% CI for $\mathbb{E}(Y_{1*})$ is $\mathbf{x}^{*T}\hat{\boldsymbol{\mu}} \pm \tilde{\sigma}\,\tau\, t_{n-p}(\frac{\alpha}{2})$

$$= 124.3 \pm 10.4/\sqrt{5} \times 2.09 = 124.3 \pm 4.66 \times 2.09 = (114.6, 134.0).$$

- Note that we are using an estimate of σ obtained from all five instruments. If we had only used the data from the first instrument, σ would be estimated as $\tilde{\sigma}_1 = \sqrt{\sum_{j=1}^{5}(y_{1,j} - \overline{y}_{1.})^2/(5-1)} = 8.74$.
- The observed 95% confidence interval for μ_1 would have been

$$\overline{y_{1.}} \pm \frac{\tilde{\sigma}_1}{\sqrt{5}}\ t_4(\frac{\alpha}{2}) = 124.3 \pm 3.91 \times 2.78 = (113.5, 135.1).$$

- The 'pooled' analysis gives a slightly narrower interval.


95% confidence intervals for means
Resistivity
90
110
130
150
1
2
3
4
5
Instrument

A 95% prediction interval for Y_{1*} at $\mathbf{x}^{*T} = (1, 0, ..., 0)$ is

$$\mathbf{x}^{*T}\hat{\boldsymbol{\mu}} \pm \tilde{\sigma}\sqrt{(\tau^2+1)}\ t_{n-p}(\tfrac{\alpha}{2}) = 124.3 \pm 10.42 \times 1.1 \times 2.07 = (100.5, 148.1).$$