STAT 135
14. Linear regression and inference

Spring 2022

Lecturer: Dr Rebecca Barter (*she/her*)
Office hours: Tu 9:30-10:30 (in person), Th 4-5pm (virtual)
Office: Evans 339

Email: rebeccabarter@berkeley.edu
Twitter: @rlbarter
GitHub: rlbarter

Inference for linear regression

Our inference goal is to learn about the “true relationship” between the covariates (x) and the response (y).

Assuming that the true relationship is: $y = \beta_0 + \beta_1 x + \epsilon$

We are interested in learning about the values of β_0 and β_1

If the true β_1 is nonzero, then we know that there is a “real” relationship between x and y

But we don’t observe the “true” β_0 and β_1, we instead observe estimates of them $\hat{\beta}_0$ and $\hat{\beta}_1$ (e.g. via LS)

Inference in the context of linear regression primarily involves conducting hypothesis tests of:

$$H_0 : \beta_1 = 0 \ \text{ versus } \ H_1 : \beta_1 \neq 0$$

Assumptions for inference for linear regression

To conduct inference in the context of linear regression, we need to make the following assumptions:

1. There is actually a linear relationship between the response and predictors, as in:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

2. (a) The errors, ϵ_i, are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

2. (b) The errors, ϵ_i, are IID $N(0, \sigma^2)$

The constant variance assumption is called **homoskedasticity** (or homos<u>c</u>edasticity)

Note that the **randomness** in the data lies in the random deviations from the "true" relationship (rather than random sampling as in our previous inference adventures)

LS estimators

Recall that the LS estimators of β_0 and β_1 are given by:

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \qquad \hat{\beta}_1 = \frac{\sum_i (y_i - \overline{y})(x_i - \overline{x})}{\sum_i (x_i - \overline{x})^2} = \frac{Cov(x, y)}{Var(x)}$$

In order to develop some hypothesis tests for these coefficients, let's first examine their expected values, their variances, and distributions!

Bias and Variance of LS estimates of β_0 and β_1

$\hat{\beta}_1$ is unbiased

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

ϵ_i are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

$$\hat{\beta}_1 = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{Cov(x,y)}{Var(x)}$$

The LS estimate, $\hat{\beta}_1$ is unbiased:

$$E[\hat{\beta}_1] = \beta_1$$

Proof:

$$E[\hat{\beta}_1] = E\left[\frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}\right]$$

$$= \frac{\sum_i (x_i - \bar{x}) E(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

And:

$$E(y_i - \bar{y}) = E[\beta_0 + \beta_1 x_i + \epsilon_i - (\beta_0 + \beta_1 \bar{x} + \bar{\epsilon})]$$

$$= \beta_1 (x_i - \bar{x}) + E[\epsilon_i - \bar{\epsilon}]$$

$$= \beta_1 (x_i - \bar{x})$$

So:

$$E[\hat{\beta}_1] = \frac{\beta_1 \sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} = \beta_1$$

Therefore $\hat{\beta}_1$ is unbiased

$\hat{\beta}_0$ is unbiased

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

ϵ_i are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

The LS estimate, $\hat{\beta}_0$ is unbiased:

$$\boxed{E[\hat{\beta}_0] = \beta_0}$$

Proof:

$$E[\hat{\beta}_0] = E\left[\bar{y} - \hat{\beta}_1 \bar{x}\right]$$

$$= E[(\beta_0 + \beta_1 \bar{x} + \bar{\epsilon})] - \beta_1 \bar{x}$$

$$= \beta_0 + \beta_1 \bar{x} + E[\bar{\epsilon}] - \beta_1 \bar{x}$$

$$= \beta_0$$

Therefore $\hat{\beta}_0$ is unbiased

Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$:

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ ϵ_i are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

The variance (and covariance) of the LS estimates $\hat{\beta}_1$ and $\hat{\beta}_0$ are given by:

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$Var(\hat{\beta}_1) = \frac{n\sigma^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

(You will prove this in homework 7)

LS estimates (β_0 and β_1) are the MLE

$\hat{\beta}_0$ and $\hat{\beta}_1$ are the MLE

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$\epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$ —(Note that for this result we **do** need to assume Normality!)

Recall the LS estimates for β_0 and β_1 are:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = \frac{Cov(x, y)}{Var(x)}$$

$$\boxed{\text{The LS estimates, } \hat{\beta}_0, \hat{\beta}_1 \text{ correspond to the MLE}}$$

(You will prove this in homework 7)

$\hat{\beta}_0$ and $\hat{\beta}_1$ are asymptotically normal

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$$

Since the LS estimates (with the normality assumption) are the MLE, this means that they are asymptotically normal:

The LS estimates, $\hat{\beta}_0, \hat{\beta}_1$ are normal

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{n\sigma^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\right)$$

Inference (hypothesis testing and confidence intervals) for β_0 and β_1

Inference for $\hat{\beta}_0$ and $\hat{\beta}_1$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0, \sigma^2)$$

Why might we want to do a hypothesis test?

Hypothesis test:

$H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$

If we find evidence against $H_0 : \beta_1 = 0$ in favor of $H_1 : \beta_1 \neq 0$, then this indicates that there a “real” relationship between x and y

Conversely if we do not find evidence against $H_0 : \beta_1 = 0$ (i.e. we accept the null), then this indicates that there is no relationship between x and y

Inference for $\hat{\beta}_0$ and $\hat{\beta}_1$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$$

If we knew $\sqrt{Var(\hat{\beta}_j)} = \sigma_{\hat{\beta}_j}$,

Hypothesis test:

$H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$

Test statistic: $\dfrac{\hat{\beta}_j - 0}{\sigma_{\hat{\beta}_j}} \sim N(0,1)$

Confidence interval:

CI: $[\hat{\beta}_j - z_{\alpha/2}\sigma_{\hat{\beta}_j}, \hat{\beta}_j + z_{\alpha/2}\sigma_{\hat{\beta}_j}]$

But do we know $\sigma_{\hat{\beta}_j}$?

The formulas are:

$$\sigma_{\hat{\beta}_0} = \sqrt{\frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{n\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

These both require that we know $Var(\epsilon_i) = \sigma^2$***… which we don’t***

Inference for $\hat{\beta}_0$ and $\hat{\beta}_1$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\boxed{\sigma^2})$$

If we have an estimate $\hat{\sigma}_{\hat{\beta}}$:

Hypothesis test:

$H_0 : \beta_j = 0$ against $H_1 : \beta_j \neq 0$

Test statistic: $\dfrac{\hat{\beta}_j - 0}{\boxed{\hat{\boldsymbol{\sigma}}_{\hat{\beta}_j}}} \sim t_{n-p}$

Confidence interval:

CI: $[\hat{\beta}_j - t_{n-p,\alpha/2}\boxed{\hat{\boldsymbol{\sigma}}_{\hat{\beta}_j}}, \hat{\beta}_j + t_{n-p,\alpha/2}\boxed{\hat{\boldsymbol{\sigma}}_{\hat{\beta}_j}}]$

Computing an estimate, $\boxed{\hat{\sigma}_{\hat{\beta}},}$ requires an estimate $\boxed{\hat{\sigma}^2}$ (of $Var(\epsilon_i) = \boxed{\sigma^2}$)

$$\boxed{\hat{\sigma}_{\hat{\beta}_0}} = \sqrt{\frac{\boxed{\hat{\sigma}^2}\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

$$\boxed{\hat{\sigma}_{\hat{\beta}_1}} = \sqrt{\frac{n\boxed{\hat{\sigma}^2}}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

But how to estimate $Var(\epsilon_i) = \boxed{\sigma^2}$ since we don't observe the ϵ_i?

When do we need Normality of ϵ?

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ ϵ_i are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

We **do not** the normality assumption for the unbiasedness and variance calculations:

$$E[\hat{\beta}_0] = \beta_0 \qquad E[\hat{\beta}_1] = \beta_1$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$Var(\hat{\beta}_1) = \frac{n\sigma^2}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}$$

When do we need Normality of ϵ?

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\boxed{\sigma^2})$$

We **do** need the normality assumption for the MLE asymptotic normality:

$$\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\right) \qquad \hat{\beta}_1 \sim N\left(\beta_1, \frac{n\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}\right)$$

On which the hypothesis testing test statistic distributional assumptions rest

Test statistic: $\dfrac{\hat{\beta}_j}{\sigma_{\hat{\beta}_j}} \sim N(0,1)$

Test statistic: $\dfrac{\hat{\beta}_j}{\hat{\boldsymbol{\sigma}}_{\hat{\beta}_j}} \sim t_{n-p}$

Estimating $E(\epsilon_i) = \sigma^2$ using the residuals

Estimating σ using the residuals

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$$

We need to estimate $Var(\epsilon_i) = \sigma^2$, but we don't observe the ϵ_i

Rearranging the linear model, the random deviations from the true line are:

$$\epsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

We don't *observe* this, but we can estimate it by plugging in $\hat{\beta}_0$ and $\hat{\beta}_1$

The **residuals** are the (training) prediction errors:

$$r_i \;=\; y_i - \hat{y}_i \;=\; y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

We can *compute* this!

Idea: The variance of the residuals is a reasonable approximation for the variance of the ϵ_is (σ^2)

Estimating σ using the residuals

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0, \sigma^2)$$

The **residuals** are the (training) error terms: $r_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Idea: The variance of the residuals is a reasonable approximation for the variance of the ϵ_is (σ^2)

Let's estimate σ^2 using the **residual sum of squares (RSS = $\sum_i r_i^2$)**:

$$\boxed{\hat{\sigma}^2 = \frac{RSS}{n-p} = \frac{1}{n-p} \sum_{i=1}^{n} r_i^2}$$

(p is the number of terms in the regression, here $p = 2$)

$\hat{\sigma}^2$ Is an unbiased estimator for σ^2

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$$

The **residuals** are the (training) prediction errors: $r_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

$\hat{\sigma}^2$ is an unbiased estimator of σ^2

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n-p}\sum_{i=1}^{n} r_i^2\right] = \sigma^2$$

(p is the number of terms in the regression, here $p = 2$)

(We will prove this in matrix form)

Hypothesis tests for β_0 and β_1

Hypothesis tests for $\hat{\beta}_0$ and $\hat{\beta}_1$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$$

$$H_0 : \beta_j = 0 \text{ against } H_1 : \beta_j \neq 0$$

Test statistic: $t = \dfrac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}}$

P-value: $P(|T| \geq |t|)$

Where $T \sim t_{n-p}$
(p is the number of parameters in the model)

Where

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{\frac{\boxed{\hat{\sigma}^2}\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{n\boxed{\hat{\sigma}^2}}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}}$$

And $\boxed{\hat{\sigma}^2} = \dfrac{RSS}{n-p} = \dfrac{1}{n-p}\displaystyle\sum_{i=1}^{n} r_i^2$

Confidence intervals for β_0 and β_1

Confidence intervals for $\hat{\beta}_0$ and $\hat{\beta}_1$

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \overset{IID}{\sim} N(0, \sigma^2)$$

$(1-\alpha)\,\%$ **Confidence interval:**

CI: $[\hat{\beta}_j - t_{n-p,\alpha/2}\hat{\sigma}_{\hat{\beta}_j}, \ \hat{\beta}_j + t_{n-p,\alpha/2}\hat{\sigma}_{\hat{\beta}_j}]$

Where

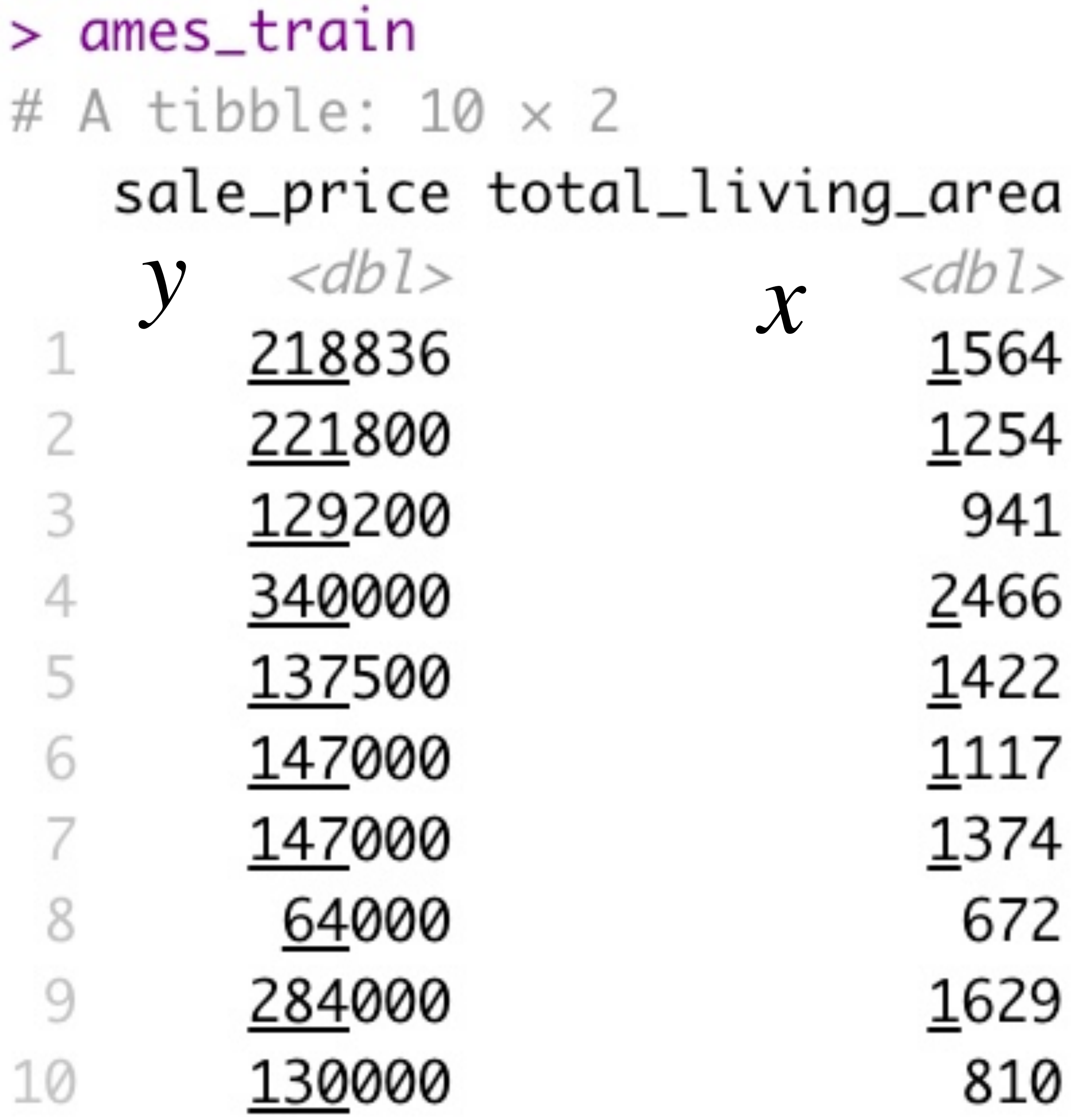
$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{\frac{\boxed{\hat{\sigma}^2}\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} \qquad \hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{n\boxed{\hat{\sigma}^2}}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}}$$

And

$$\boxed{\hat{\sigma}^2} = \frac{RSS}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n} r_i^2$$

Toy example

$$\widehat{sale\_price} = \hat{\beta}_0 + \hat{\beta}_1 area$$



```
> ames_train
# A tibble: 10 × 2
   sale_price total_living_area
        <dbl>             <dbl>
 1     218836              1564
 2     221800              1254
 3     129200               941
 4     340000              2466
 5     137500              1422
 6     147000              1117
 7     147000              1374
 8      64000               672
 9     284000              1629
10     130000               810
```

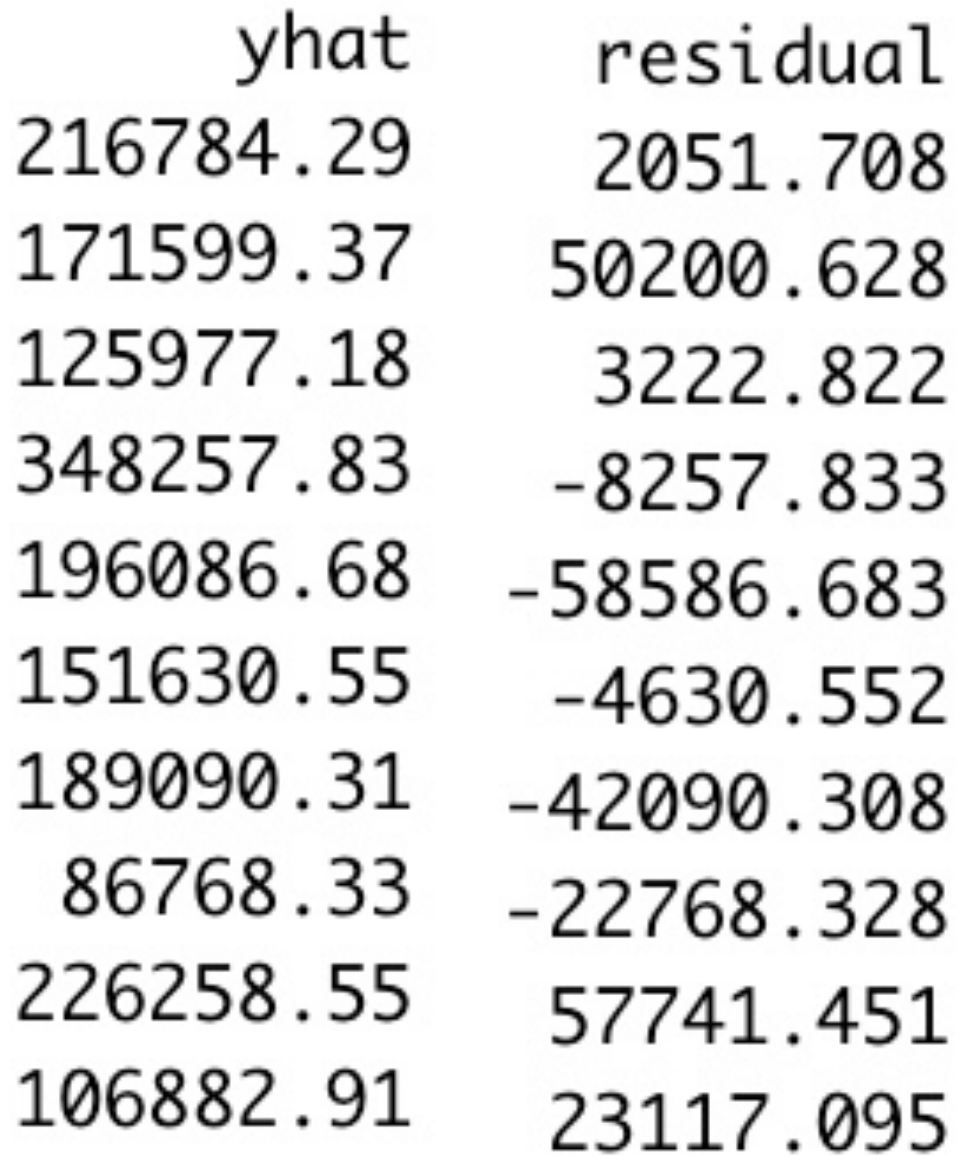
y x

$$\bar{x} = 1324.9$$

$$\bar{y} = 181933.6$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} r_i^2 = 1{,}526{,}899{,}810$$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ $\quad r$



```
     yhat    residual
216784.29    2051.708
171599.37   50200.628
125977.18    3222.822
348257.83   -8257.833
196086.68  -58586.683
151630.55   -4630.552
189090.31  -42090.308
 86768.33  -22768.328
226258.55   57741.451
106882.91   23117.095
```

$$H_0 : \beta_0 = 0 \qquad H_1 : \beta_0 \neq 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -11180.92$$

$$\hat{\sigma}_{\hat{\beta}_0} = \sqrt{\frac{\hat{\sigma}^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = 35{,}957.24$$

$$t_{\hat{\beta}_0} = \hat{\beta}_0 / \hat{\sigma}_{\hat{\beta}_0} = -0.31 \qquad P(|T| \geq 0.31) = 0.76$$

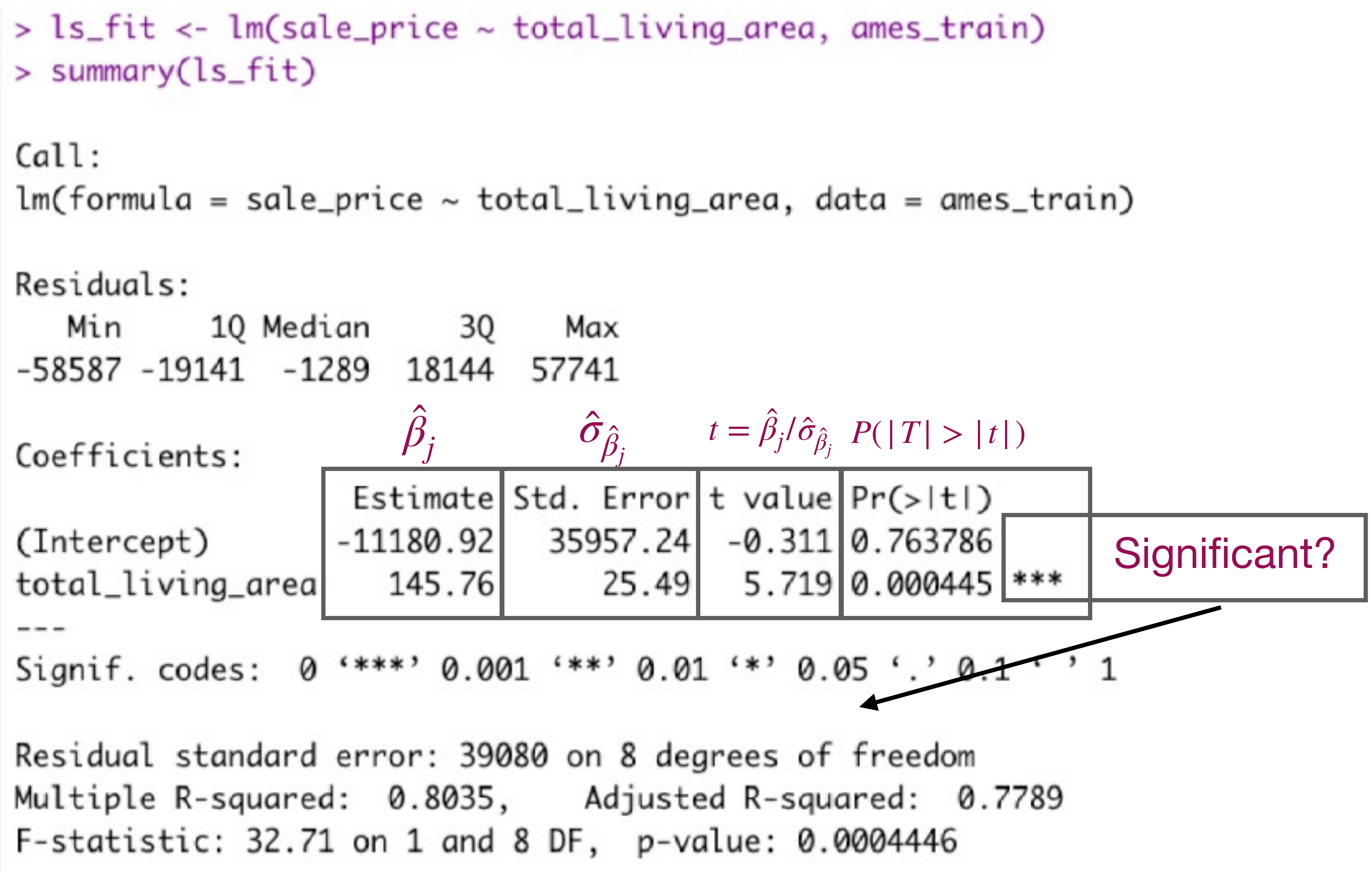
$$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$$

$$\hat{\beta}_1 = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} = 145.76$$

$$\hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{n\hat{\sigma}^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = 25.49$$

$$t_{\hat{\beta}_1} = \hat{\beta}_1 / \hat{\sigma}_{\hat{\beta}_1} = 5.72 \qquad P(|T| \geq 5.72) = 0.0044$$

Example



```
> ls_fit <- lm(sale_price ~ total_living_area, ames_train)
> summary(ls_fit)

Call:
lm(formula = sale_price ~ total_living_area, data = ames_train)

Residuals:
   Min     1Q Median     3Q    Max 
-58587 -19141  -1289  18144  57741 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       -11180.92   35957.24  -0.311 0.763786    
total_living_area    145.76      25.49   5.719 0.000445 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39080 on 8 degrees of freedom
Multiple R-squared:  0.8035,	Adjusted R-squared:  0.7789 
F-statistic: 32.71 on 1 and 8 DF,  p-value: 0.0004446
```

$\hat{\beta}_j$

$\hat{\sigma}_{\hat{\beta}_j}$

$t = \hat{\beta}_j / \hat{\sigma}_{\hat{\beta}_j}$

$P(|T| > |t|)$

Significant?

Example

lm_inference.R

Residual plots for assessing inference assumptions

Assumptions for inference for linear regression

To conduct inference in the context of linear regression, we need to make the following assumptions:

1. There is actually a linear relationship between the response and predictors, as in:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

2. (a) The errors, ϵ_i, are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

2. (b) The errors, ϵ_i, are IID $N(0,\sigma^2)$

The constant variance assumption is called **homoskedasticity** (or homos<u>c</u>edasticity)

Evaluating homoskedasticity

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \boxed{Var(\epsilon_i) = \sigma^2}$$

Let's talk about this assumption

Homoskedasticity:
The variance of the error associated with each observation is identical and does not depend on x

Heteroskedasticity:
The variance of the error associated with each observation is different and may depend on x

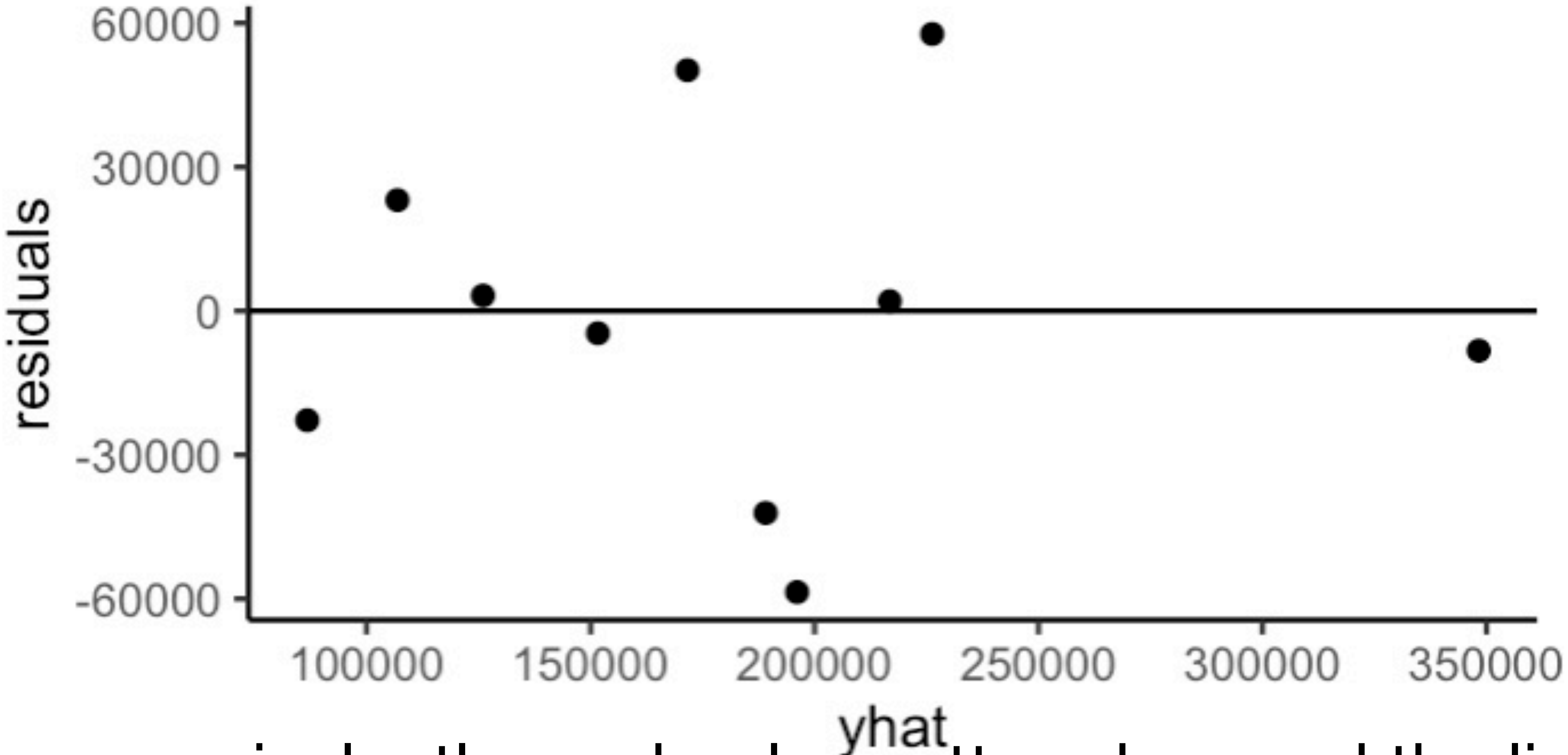
We don't observe ϵ, but perhaps we can get a sense for whether these assumptions are reasonable using the residuals

$$r_i = y_i - \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Visualizing predictive performance: residual plot

A residual plot plots the residuals $r_i = y_i - \hat{y}_i$ against the fitted values, $\hat{y}_i$ (or covariate values x_i)

$$\widehat{sale\_price} = \hat{\beta}_0 + \hat{\beta}_1 area$$

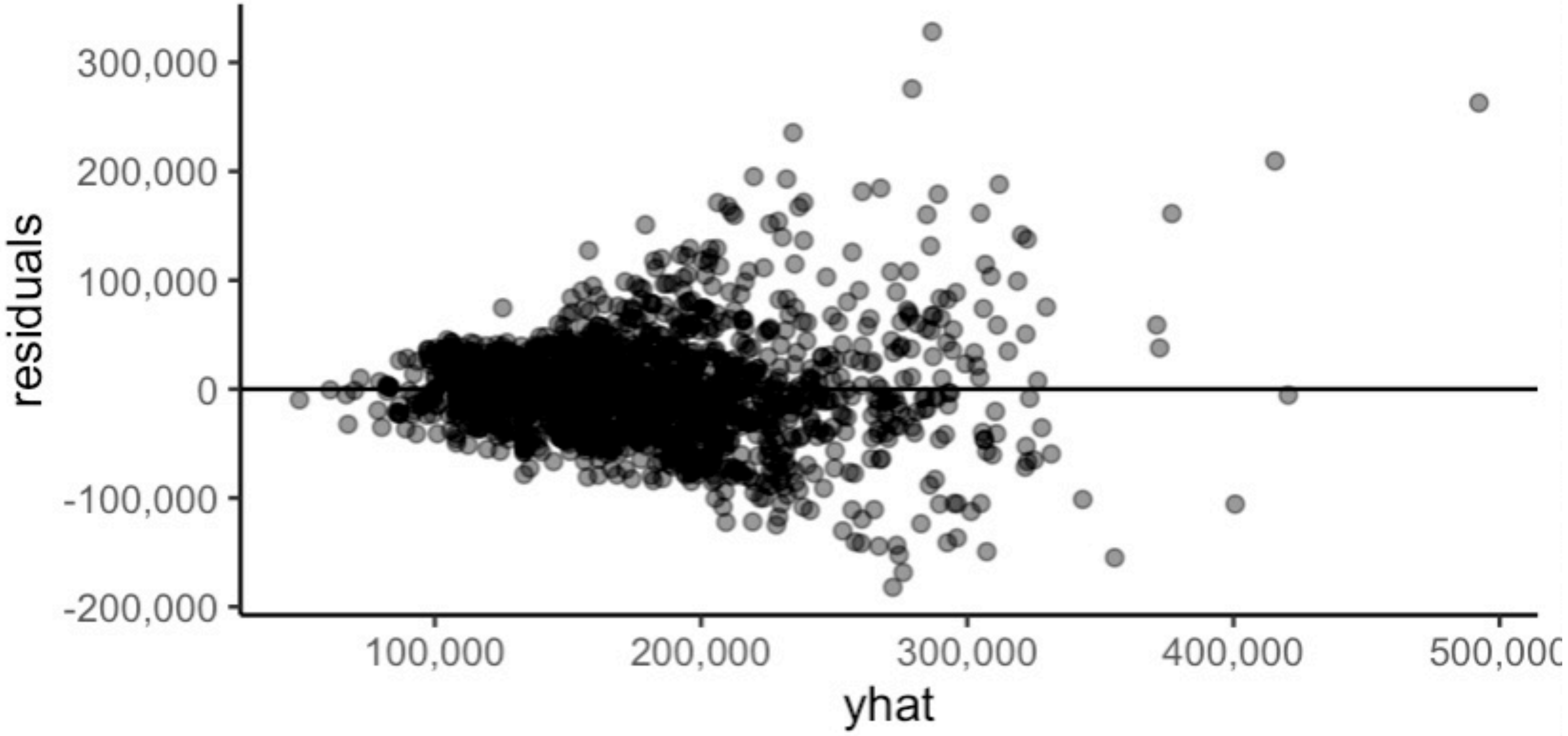


If the points seem equivalently randomly scattered around the line y=0, then this is evidence that the linear relationship and the homoskedasticity assumptions are satisfied

Visualizing predictive performance: residual plot

A residual plot plots the residuals $r_i = y_i - \hat{y}_i$ against the fitted values, $\hat{y}_i$ (or covariate values x_i)

$$\widehat{sale\_price} = \hat{\beta}_0 + \hat{\beta}_1 area$$



The residuals are **more variable** for larger predicted response values — **heteroskedastic**

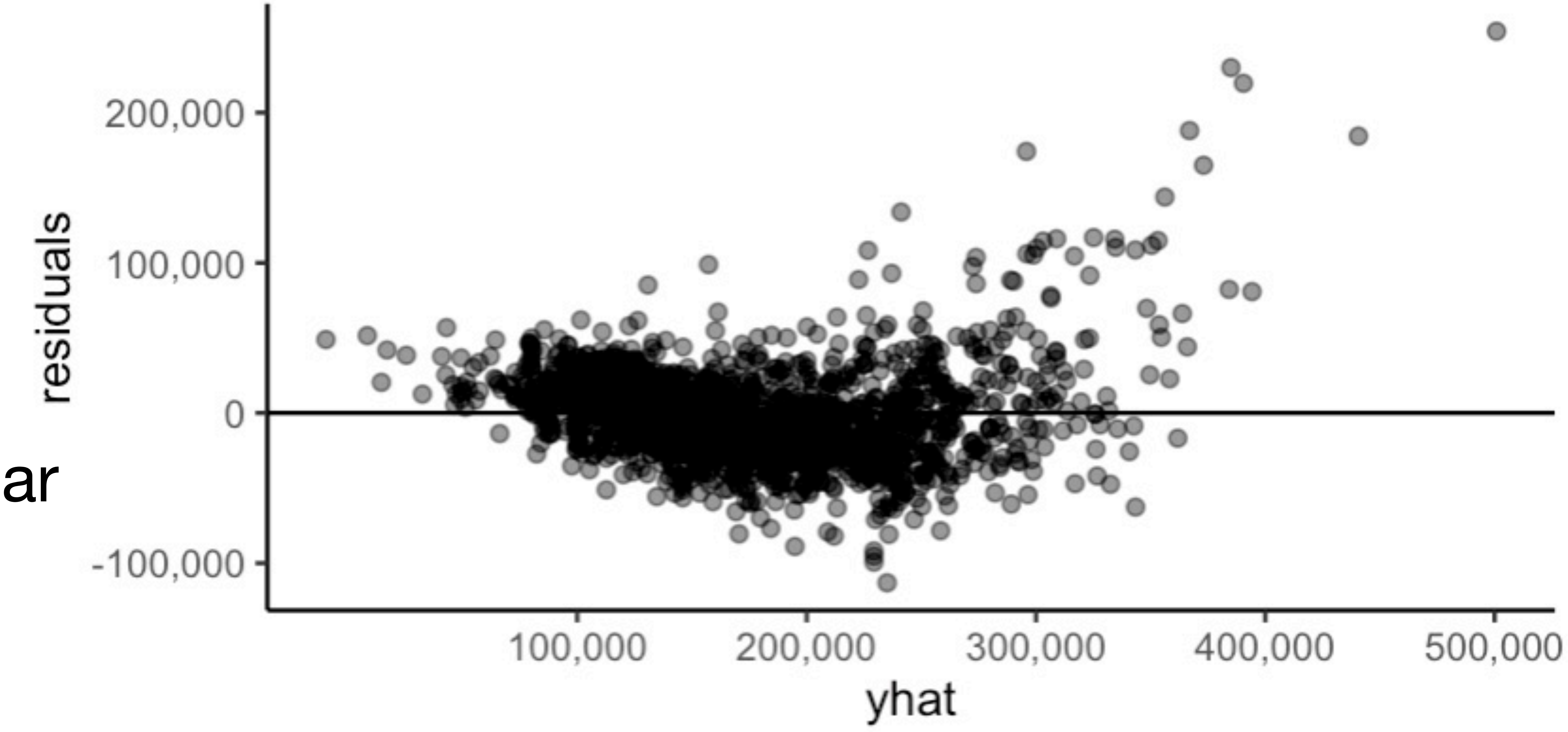
Visualizing predictive performance: residual plot

A residual plot plots the residuals $r_i = y_i - \hat{y}_i$ against the fitted values, $\hat{y}_i$ (or covariate values x_i)

$$\widehat{sale\_price} = \hat{\beta}_0 + \hat{\beta}_1 area + \hat{\beta}_2 bedrooms + \hat{\beta}_3 quality + \hat{\beta}_4 year$$

The more complex fit is better, but the residuals still don't look totally randomly distributed around 0

This implies that either (1) the linear relationship or (2) the common variance assumption are not completely satisfied… so any inference conclusions should be taken with a grain of salt…



Example

lm_inference.R