Stat 23400 Lecture 17: Statistical Inference of Simple Linear Regression Models

Outline for Today

- Topics:
  - Introduction to Simple Linear Regression (MMSA 12.1, 12.2 and OS4 8.1, 8.2)
  - Statistical Inference for Simple Linear Regression (MMSA 12.3, 12.4; OS4 8.4).

- Hw9 will be posted on Canvas.

Introduction to Simple Linear Regression Models

Introduction to Linear Models I

response (random)		mean depends on another variable (x) (not random)		error (random)
$\downarrow$		$\downarrow$		$\downarrow$
Y	$=$	$\mu(x)$	$+$	ϵ

- We consider the x-values as "fixed" and model the probability distribution of Y "conditional" on the observed x-values. Y is referred to as **dependent variable** (or response) and x is as **independent variable** (or **predictor**).

Simple Linear Regression Model

$$\underset{\substack{\text{response} \\ \text{(random)}}}{\overset{}{Y}} = \underset{\substack{\text{mean depends} \\ \text{on another} \\ \text{variable } (x) \\ \text{(not random)}}}{\mu(x)} + \underset{\substack{\text{error} \\ \text{(random)}}}{\epsilon}$$

response (random) $\downarrow$ Y = mean depends on another variable (x) (not random) $\downarrow$ $\mu(x)$ + error (random) $\downarrow$ ϵ

- The mean is a linear function of x :

$$\mu(x) = \mu_{Y|x} = E(Y|x) = \beta_0 + \beta_1 x.$$

- **Error** or **noise** term:

$$Y_i - (\beta_0 + \beta_1 x_i) = \epsilon_i\,, \qquad \epsilon_i \sim \text{ i.i.d. N}(0, \sigma^2) \quad i = 1, \ldots, n\,.$$

- This model is called the **simple linear regression model**.

The linear model is

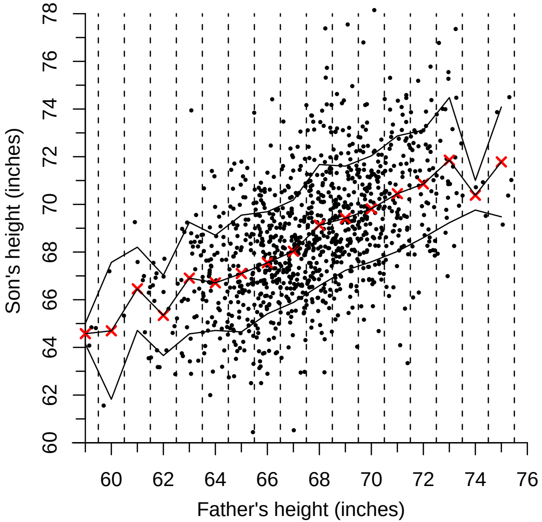
$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i\,, \qquad \epsilon_i \sim \text{ i.i.d. N}(0, \sigma^2) \quad i = 1, \ldots, n$$

- We observe pairs (x_i, y_i).
- What are the expected value and variance of Y_i given x_i?

- **assumption:** constant variance for Y for every value of predictor variable (x).

Example: Pearson's Father-and-Son Data

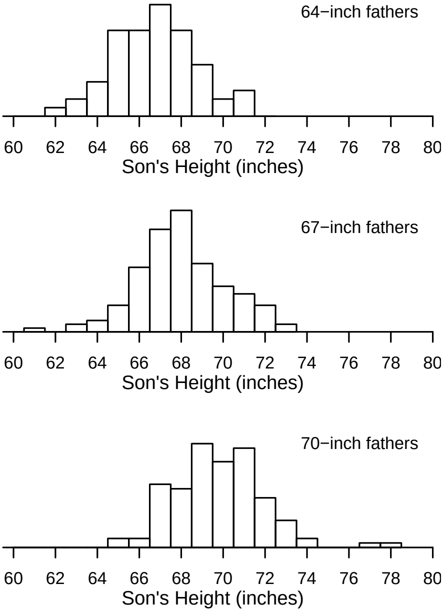
Father-son pairs are grouped by father's height, to the nearest inch.



How do the

- mean of son's height (SH),
- SD of SH, and
- distribution of SH (histogram of SH)

within each group change with father's height (FH)?


64−inch fathers
60 62 64 66 68 70 72 74 76 78 80
Son's Height (inches)
67−inch fathers
60 62 64 66 68 70 72 74 76 78 80
Son's Height (inches)
70−inch fathers
60 62 64 66 68 70 72 74 76 78 80
Son's Height (inches)

Simple Linear Regression Model

Pearson's father-and-son data inspire the following assumptions for the simple linear regression (SLR) model:

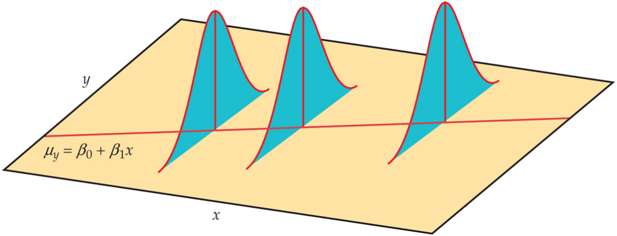
1. The means of Y is a linear function of X, i.e.,

$$E(Y \mid X = x) = \beta_0 + \beta_1 x$$

2. The SD of Y does not change with x, i.e.,

$$SD(Y \mid X = x) = \sigma \quad \text{for every } x$$

3. (Optional) Within each subpopulation, the distribution of Y is normal.



Statistical Inference of Simple Linear Regression Models

Data for a Simple Linear Regression Model

- Suppose we have a SRS of n individuals from a population. From individual i we observe the response y_i and the explanatory variable x_i:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$

.

- The SLR model states that

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

- To estimate the regression model, we can use the least square line of the data.

$$y = b_0 + b_1 x$$

in which

$$b_1 = r\frac{s_y}{s_x} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}, \quad b_0 = \overline{y} - b_1 \overline{x}$$

Caution: Sample v.s. Population

Note the population regression line

$$y = \beta_0 + \beta_1 x$$

is *different* from the least square regression line

$$y = b_0 + b_1 x$$

.

- The LS line is for a sample, while population regression line is for the entire population.
- b_0 and b_1 will change from sample to sample.

$$b_1 = r\frac{s_y}{s_x} = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}, \quad b_0 = \overline{y} - b_1\overline{x}$$

- Interest in β_0 and β_1, NOT the sample counterparts b_0 and b_1.

How Close Are b_0, b_1 to β_0, β_1?

Recall LS estimate is

$$b_1 = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2}, \quad b_0 = \overline{y} - b_1 \overline{x}.$$

We can show that,

$$b_1 = \beta_1 + \frac{\sum_i (x_i - \overline{x})\varepsilon_i}{\sum_i (x_i - \overline{x})^2}, \qquad b_0 = \beta_0 + \overline{\varepsilon} - \frac{\sum_i \overline{x}(x_i - \overline{x})\varepsilon_i}{\sum_i (x_i - \overline{x})^2}.$$

From the above, we have

$$E(b_1) = \beta_1 \text{ and } E(b_0) = \beta_0$$

b_0, b_1 are **unbiased** estimates.

Variability of b_1

- One can show that

$$SD(b_1) = \frac{\sigma}{\sqrt{\sum (x_i - \overline{x})^2}} = \frac{\sigma}{s_x \sqrt{n-1}},$$

where $s_x = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n-1}}$ is the sample SD of x_i's.

- How to reduce the SD of b_1 (and make b_1 closer to β_1):
  - increase the sample size n
  - increase the range of x_i's (and hence s_x is increased)
- But σ is unknown, we need to estimate it.

Estimate of σ

We want to estimate σ, SD of the error ε_i.

- An intuitive estimate of σ is the sample SD of the errors ε_i

$$\widehat{\sigma} = \sqrt{\frac{\sum(\varepsilon_i - \overline{\varepsilon})^2}{n-1}} \quad \text{where} \quad \varepsilon_i = y_i - \beta_0 - \beta_1 x_i$$

- However, this is not possible β_0 and β_1 are unknown.
- We can estimate β_0 and β_1 with b_0 and b_1 and approximate the errors ε_i with the **residuals**

$$e_i = y_i - (b_0 + b_1 x_i) = y_i - \widehat{y}_i$$

- We use the "sample SD" of the residuals e_i to estimate σ:

$$s_e = \sqrt{\frac{\sum(e_i - \overline{e})^2}{n-2}} = \sqrt{\frac{\sum e_i^2}{n-2}}$$

  - Recall that the mean of residuals is 0, $\overline{e} = \sum_i e_i / n = 0$.
  - Note here we divide by $n-2$, not $n-1$.

Sampling distribution of b_1

Recall that

$$b_1 = \beta_1 + \frac{\sum_i (x_i - \overline{x})\varepsilon_i}{\sum_i (x_i - \overline{x})^2}$$

- The **sampling distribution** of b_1 is normal

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum (x_i - \overline{x})^2}\right) \Rightarrow Z = \frac{b_1 - \beta_1}{\sigma / \sqrt{\sum (x_i - \overline{x})^2}} \sim N(0,1)$$

- But σ is unknown, we estimate it with s_e. The estimated SD of b_1 is called the **standard error (SE)** of b_1

$$SE(b_1) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \overline{x})^2}} = \frac{s_e}{\sqrt{\sum (x_i - \overline{x})^2}}\,.$$

- This leads to t-statistic, with $n-2$ degrees of freedom

$$T = \frac{b_1 - \beta_1}{s_e / \sqrt{\sum (x_i - \overline{x})^2}} = \frac{b_1 - \beta_1}{SE(b_1)} \sim t_{n-2}\,.$$

Find a $(1-\alpha)$ confidence interval for β_1 based on

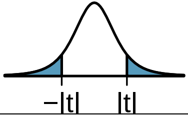
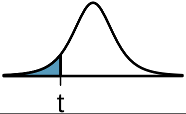
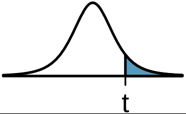
$$T = \frac{b_1 - \beta_1}{s_e/\sqrt{\sum(x_i - \overline{x})^2}} = \frac{b_1 - \beta_1}{SE(b_1)} \sim t_{n-2}\,.$$

Tests for β_1

To **test the hypothesis** $H_0 : \beta_1 = a$, we use the t-statistic

$$T = \frac{b_1 - a}{SE(b_1)} \sim t_{n-2}$$

The p-value can be computed using the t-table based on the H_a:

H_a	$\beta_1 \neq a$	$\beta_1 < a$	$\beta_1 > a$
P-value			

Observe that testing $H_0 : \beta_1 = 0$ is equivalent to testing whether x is useful in predicting y linearly.

- It is possible that r is small but β_1 is significantly different from 0.

Inference for the Intercept β_0

Though **the population intercept β_0 is rarely of interest**, all the results for the population slope β_1 have their counterparts for β_0.

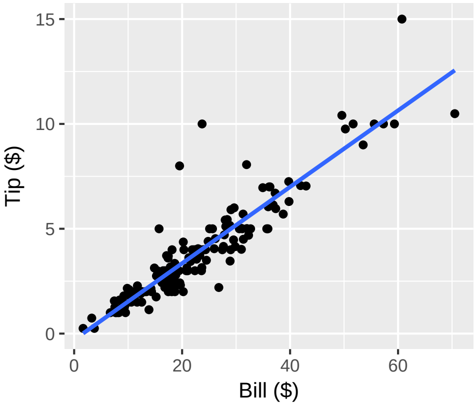
- $b_0 = \beta_0 + \bar{\varepsilon} - \frac{\sum_i \bar{x}(x_i - \bar{x})\varepsilon_i}{\sum_i (x_i - \bar{x})^2}$.
- $E(b_0) = \beta_0$ . . . . (b_0 is an **unbiased** estimate of β_0).
- $\mathrm{SD}(b_0) = \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$.
- $\mathrm{SE}(b_0) = s_e\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$.
- The sampling distribution of b_0 (when n is large) is

$$b_0 \sim N\left(\beta_0, \left(\sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}\right)^2\right)$$

- $(1-\alpha)$ C.I. for β_0: $b_0 \pm t_{n-2,\alpha/2} SE(b_0)$.
- The test statistic for $H_0 : \beta_0 = a$ is $t = \dfrac{b_0 - a}{SE(b_0)} \sim t_{n-2}$ and the P-value can be computed similarly as for β_1

Example: Restaurant Tips

The owner of a bistro called *First Crush* in Potsdam, NY, collected 157 restaurant bills over a 2-week period that he believes provide a good sample of his customers. He wanted to study the payment and tipping patterns of its patrons.



Regression in R is as simple as `lm(y ~ x)`, in which "`lm`" stands for "`l`inear `m`odel"

```
> tips = read.table("RestaurantTips.txt",h=T)
> lm(Tip ~ Bill, data=tips)

Call:
lm(formula = Tip ~ Bill, data = tips)

Coefficients:
(Intercept)         Bill
    -0.2923       0.1822
```

It is better to save the model as an object,

```
lmtips = lm(Tip ~ Bill, data=tips)
```

and then we can get a more detailed output by viewing the `summary()` of the model object. The output is shown in the next slide

Regression in R

```
> summary(lmtips)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.292267   0.166160  -1.759   0.0806 .
Bill         0.182215   0.006451  28.247   <2e-16 ***
---
Residual standard error: 0.9795 on 155 degrees of freedom
Multiple R-squared:  0.8373,Adjusted R-squared:  0.8363
F-statistic: 797.9 on 1 and 155 DF,  p-value: < 2.2e-16
```

- LS estimate for the intercept b_0 and the slope b_1.
- The "`Std. Error`" $\mathrm{SE}(b_0)$ and $\mathrm{SE}(b_1)$.
- t-values $b_i/SE(b_i)$ are the ratio of "`Estimate`" and "`Std. Error`", e.g.,

$$-1.759 = \frac{-0.292267}{0.166160}, \quad 28.247 = \frac{0.182215}{0.006451}$$

- The small P-value $< 2 \times 10^{-16}$ asserts that the amount of tips is linearly related to the amount of the bill significantly.

Example: Confidence Interval for β_1

How to get CI for β_1?

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.292267  0.166160  -1.759  0.0806
Bill         0.182215  0.006451  28.247  <2e-16
```

As df $= n - 2 = 157 - 2 = 155$, $t_{155,0.05/2}$ for a 95% CI is 1.975 (between 1.97 and 1.98).

one tail two tails	0.1 0.2	0.05 0.10	0.025 0.050	0.01 0.02	0.005 0.010
df 150	1.29	1.66	1.98	2.35	2.61
200	1.29	1.65	1.97	2.35	2.60

Hence the 95% CI for β_1 is

$$b_1 \pm t_{155,0.05/2} SE(b_1) = 0.182215 \pm 1.975 \times 0.006451$$
$$= 0.182215 \pm 0.01274 \approx (0.169, 0.195).$$

Interpretation: With 95% confidence, for each additional dollar in the bill, the customers gave 16.9 cents to 19.5 cents more tips on average.

Example: Test for the Slope β_1

A general rule for waiters is to tip 15 to 20% of the pre-tax bill. That is, $\beta_0 = 0$ and β_1 is between 0.15 to 0.20.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.292267	0.166160	-1.759	0.0806
Bill	0.182215	0.006451	28.247	<2e-16

- R tests
  - $\beta_0 = 0$: t-statistic $= -1.759$, 2-sided p-value $= 0.0806$.
  - $\beta_1 = 0$: t-statistic $= 28.247$, 2-sided p-value $< 2e - 16$.
- How to test $H_0 : \beta_1 = 0.2$ v.s. $H_A : \beta_1 < 0.2$?

How to Read R Outputs for Regression?

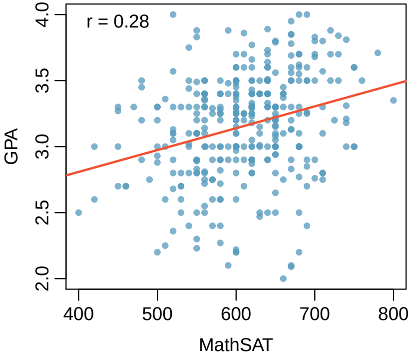
```
Residual standard error: 0.9795 on 155 degrees of freedom
Multiple R-squared:  0.8373,Adjusted R-squared:  0.8363
F-statistic: 797.9 on 1 and 155 DF,  p-value: < 2.2e-16
```

- `Residual standard error: 0.9795 on 155 degrees of freedom`
  This gives the estimate s_e of σ, which is 0.9795.
  $\text{df} = n - 2 = 157 - 2 = 155$
- `Multiple R-squared: 0.8373` gives $r^2 = 0.8373$, Bill size explained 83.73% of the variation in tipping amount.
  The correlation between bill size and tips is
  $r = \sqrt{r^2} = \sqrt{0.8373} = 0.915$.
- `Adjusted R-squared`: Ignore this.
- `F-statistic: 797.9 on 1 and 155 DF,  p-value: < 2.2e-16`
  Skip.

More Examples

Example: GPA and MathSAT

The scatter plot below shows the GPA and MathSAT of a random sample of 345 students in a college.



The correlation $r = 0.28$ is weak.

Question: Can the slope β_1 of the line be significantly different from 0?

Example: GPA and MathSAT

```
> summary(lm(GPA ~ VerbalSAT, data=stu))
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.1466877  0.1867166  11.497  < 2e-16 ***
MathSAT     0.0016544  0.0003036   5.449 9.68e-08 ***
```

To test $H_0 : \beta_1 = 0$ v.s. $H_1 : \beta_1 \neq 0$, the t-statistic is

$$t = \frac{b_1}{SE(b_1)} = \frac{0.0016544}{0.0003036} = 5.449$$

with df $= 345 - 2 = 343$. Two-sided P-value $= 9.68 \times 10^{-8}$.

- There is strong evidence that students' GPA is linearly related with their MathSAT, despite of their small correlation $r = 0.28$.
- It is possible that r is small but β_1 is significantly different from 0, especially when the sample size n is large.
- Students with higher MathSAT indeed have significantly higher GPA on average, despite of the huge variability in GPA.
- As $R^2 = r^2 = (0.28)^2 = 0.0784$, MathSAT merely explains 7.84% of the variation in GPA.

Example: GPA and MathSAT – 95% CI for β_1

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.1466877  0.1867166  11.497  < 2e-16 ***
MathSAT     0.0016544  0.0003036   5.449 9.68e-08 ***
```

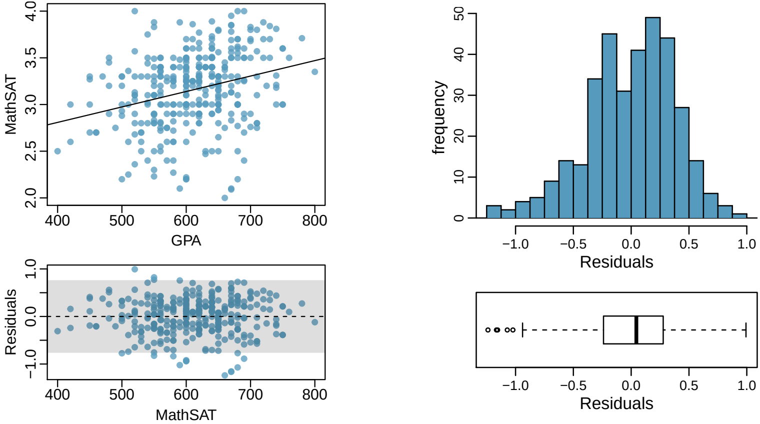
df $= 345 - 2 = 343$. The t^* for a 95% CI is 1.97.

one tail two tails	0.1 0.2	0.05 0.10	0.025 0.050	0.01 0.02	0.005 0.010
df 300	1.28	1.65	1.97	2.34	2.59
400	1.28	1.65	1.97	2.34	2.59

So a 95% confidence interval for β_1 is

$$b_1 \pm t_{343,0.05/2} SE(b_1) = 0.0016544 \pm 1.97 \times 0.0003036 \approx (0.00106, 0.00225$$

Interpretation: We have 95% confidence that for students with 100 more points in their MathSAT scores, their GPA are 0.106 to 0.225 higher on average.



- The linearity and constant variability conditions are fine.
- The slight left-skewness of residuals is fine because of the large sample size.

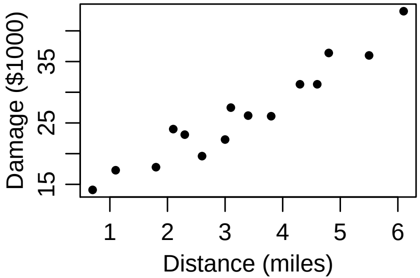
Example: Fire Damage and Distance to Fire Station

A fire insurance company wanted to relate the amount of fire damage in major residential fires to the distance between the burning house and the nearest fire station.

- The study was conducted in a large suburb of a major city; a sample of 15 recent fires in this suburb was selected.
- The amount of damage and the distance between the fire and the nearest fire station were recorded in each fire.

Fire Damage Example: Data and Scatter Plot

Distance (mile)	Damage ($1000)
0.7	14.1
1.1	17.3
1.8	17.8
2.1	24.0
2.3	23.1
2.6	19.6
3.0	22.3
3.1	27.5
3.4	26.2
3.8	26.1
4.3	31.3
4.6	31.3
4.8	36.4
5.5	36.0
6.1	43.2



Fire Damage Example: Analysis in R

```
> fire = read.table("fire.txt",h=T)
> summary(lm(damage ~ dist, data=fire))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2779     1.4203   7.237 6.59e-06 ***
dist          4.9193     0.3927  12.525 1.25e-08 ***
```

- estimate for the intercept $b_0 = 10.2779$ and the slope $b_1 = 4.9193$
- $SE(b_0) = 1.4203$, $SE(b_1) = 0.3927$

two tails	0.2	0.10	0.050	0.02	0.010
df 13	1.35	1.77	2.16	2.65	3.011

So a 95% confidence interval for β_1 is

$$b_1 \pm t_{13,0.05/2} SE(b_1) = 4.9193 \pm 2.16 \times 0.3927 \approx 4.919 \pm 0.848 \approx (4.071, 5.767).$$

Interpretation: We have 95% confidence that every extra mile from the nearest fire station increases the amount of damage by $4071 to $5767.

Example: Test for the Slope β_1

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2779     1.4203   7.237 6.59e-06 ***
dist          4.9193     0.3927  12.525 1.25e-08 ***
```

To test $H_0 : \beta_1 = 4$ v.s. $H_1 : \beta_1 > 4$, the t-statistic is

$$t = \frac{b_1 - 4}{SE(b_1)} = \frac{4.9193 - 4}{0.3927} = 2.3409$$

Looking at the t-table for the row with df $= 13$, the one-sided P-value is between 0.01 and 0.025.

one tail	0.1	0.05	0.025	0.01	0.005
two tails	0.2	0.10	0.050	0.02	0.010
df 13	1.35	1.77	2.16	2.65	3.011

Conclusion: With 95% confidence, the extra amount of damage for every extra mile from the nearest fire station is significantly higher than \$4000.

Prediction for Mean Response

Inference about Prediction

Recall the fitted line is:

$$\text{Tip} = -0.29 + 0.18\,\text{Bill}\,.$$

- How do we predict the **mean amount of tip** for Bills \$100?
- What if we want to predict **the amount of tips of a new customer** whose bill is \$100?
- How about the standard errors?

  Individual customers whose bill are \$100 don't give the same amount of tips, so the prediction for individual amount of tips has larger standard error than the prediction for mean amount of tips.

- Under SLR, what is $\mu(x^*) = E(Y \mid x = x^*)$?

  $y \sim N(\mu(x^*) = \beta_0 + \beta_1 x^*, \sigma^2)$.
- How to estimate $\mu(x^*)$?

  $\hat{\mu}(x^*) = b_0 + b_1 x^*$.
- Find a $(1-\alpha)$ confidence interval for $\mu(x^*)$.

$$\hat{\mu}(x^*) \pm t^* \mathrm{SE}(\hat{\mu}(x^*)) \, ,$$

where t^* is the upper $\alpha/2$ critical value of the t_{n-2} distribution and

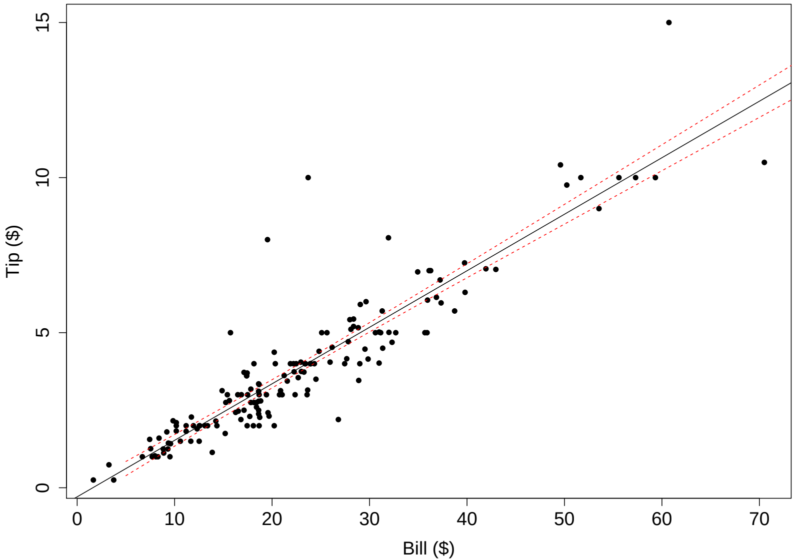
$$\mathrm{SE}(\hat{\mu}(x^*)) = s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \, .$$

Note the margin of error gets larger if x^* is farther away from $\bar{x}$.

R Analysis for Mean Response Prediction

```
newbill<-seq(5,80, by = 5)
# Get predictions for new values in newdist
prd<-predict.lm(lmtips,newdata=data.frame(Bill=newbill),
             interval = c("confidence"),
             level = 0.90,type="response")
plot(tips$Bill, tips$Tip,pch=16,xlab="Bill ($)",
     ylab="Tip ($)")
abline(lmtips) # Regression line
# Confidence bounds for prediction
lines(newbill,prd[,2],col="red",lty=2)
lines(newdist,prd[,3],col="red",lty=2)
```

Mean Response Prediction: Result



Prediction for a Future Observation(Optional)

Prediction for a Future Observation

Suppose we want to predict a specific observation value at $x = x^*$, i.e., $Y \mid x = x^*$.

- What is the distribution of $Y \mid x = x^*$?

  $Y \mid x = x^* \sim N(\mu(x^*), \sigma^2)$.

- What is the point estimate for Y?

  $\hat{y} = \hat{\mu}(x^*) = b_0 + b_1 x^*$.

- How accurate is this estimate? The error here will be larger than the error for the mean response, $\mathrm{SE}(\hat{\mu}(x^*))$, because there is error in estimating $\mu(x^*)$ as well as error in drawing a value from the normal distribution $N(\mu(x^*), \sigma^2)$.

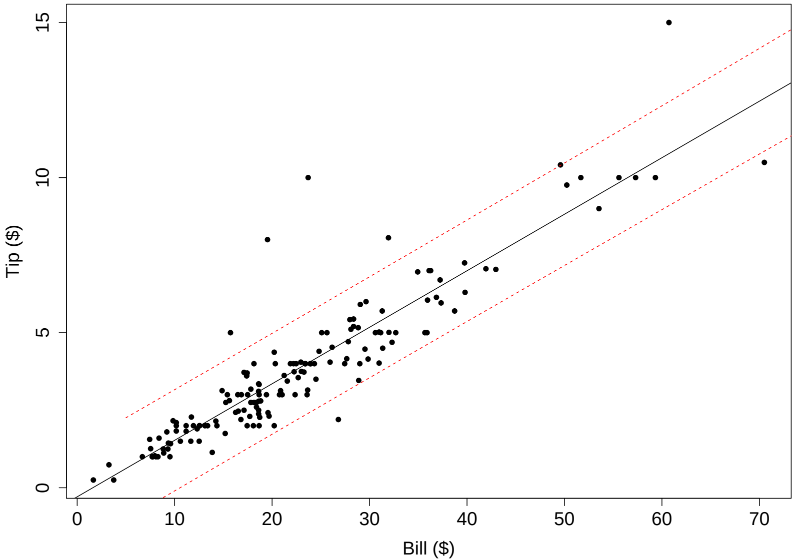
- How to find a **level** $(1-\alpha)$ **prediction interval** $Y \mid x = x^*$?

  $\hat{y} \pm t^* s_{\hat{y}}$, where t^* is the upper $\alpha/2$ critical value of the t_{n-2} distribution and $s_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^*-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$. Note again the margin of error gets larger if x^* is farther away from $\bar{x}$.

R Analysis for Predictive Interval

```r
newbill<-seq(5,80, by = 5)
# Get predictions for new values in newdist
prd<-predict.lm(lmtips,newdata=data.frame(Bill=newbill),
             interval = c("prediction"),
             level = 0.90,type="response")
plot(tips$Bill, tips$Tip,pch=16,xlab="Bill ($)",
     ylab="Tip ($)")
abline(lmtips) # Regression line
# predictive intervals
lines(newbill,prd[,2],col="red",lty=2)
lines(newdist,prd[,3],col="red",lty=2)
```

Predictive Interval: Result



Analysis of Variance (Optional)

Analysis of variance

Analysis of variance: to break down the total variation (**SST**) in data into different sources of variation.

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \,.$$

- In the regression setting, the observed variation in the responses comes from two sources.
  - **SSR: regression sum of squares** is the variation along the line as x changes.

$$\text{SSR} = \text{SSM} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \,.$$

  - **SSE: error (residual) sum of squares** is the variation for a fixed x.

$$\text{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \,.$$

The ANOVA Equation

- In LM, one can show that

$$\underbrace{\sum_{i=1}^{n}(y_i - \bar{y})^2}_{\text{SST}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}_{\text{SSR}} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\text{SSE}}$$

- The degrees of freedom break down in a similar manner:

$$\underbrace{n-1}_{\text{SST}} = \underbrace{1}_{\text{SSR}} + \underbrace{n-2}_{\text{SSE}}$$

- Mean square (MS):

$$\text{MSE} = \frac{\text{SSE}}{\text{df(Error)}} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = s^2 \, .$$

- R^2 : measures the strength of a linear fit.

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\sum_i(\hat{y}_i - \bar{y})^2}{\text{SST}} = \frac{\hat{b}_1^2 S_x^2}{S_y^2} = r^2$$

The ANOVA F Statistic

- An alternative approach to test the hypothesis: $H_0 : \beta_1 = 0$ is to use the F statistic:

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{SSR}/\text{dfR}}{\text{SSE}/\text{dfE}} = \frac{b_1^2 S_x^2}{s^2}$$
$$= \left(\frac{b_1}{s/\sqrt{S_x^2}} \right)^2 = \left(\frac{b_1}{SE(\hat{\beta}_1)} \right)^2 = t^2$$

- Under H_0,

$$F \sim F_{1,n-2}$$

where $F_{1,n-2}$ is an F distribution with 1 and $n-2$ degrees of freedom.

```
> summary(lmtips)
Call:
lm(formula = Tip ~ Bill, data = tips)
Residuals:
    Min      1Q  Median      3Q     Max
-2.3911 -0.4891 -0.1108  0.2839  5.9738
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.292267   0.166160  -1.759   0.0806 .
Bill         0.182215   0.006451  28.247   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
Residual standard error: 0.9795 on 155 degrees of freedom
Multiple R-squared:  0.8373,Adjusted R-squared:  0.8363
F-statistic: 797.9 on 1 and 155 DF,  p-value: < 2.2e-16
```

Residual standard error is the estimate s of σ. Multiple R-squared is $\frac{SSR}{SST}$. F-statistic is the square of the t-statistic for the slope.