Homework 9
Statistics 200B
Due Apr 25, 2019

1. Following the notation from class, define $RSS = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$, $ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$, and $TSS = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$. Show that $TSS = ESS + RSS$.

   Proof:

$$\begin{aligned} TSS &= \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left[(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})\right]^2 \\ &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) \\ \left(\hat{\epsilon}_i = Y_i - \hat{Y}_i\right) &= RSS + ESS + 2\sum_{i=1}^{n}\hat{\epsilon}_i(\hat{Y}_i - \bar{Y}). \end{aligned}$$

   In the simple linear regression model fitting,

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i\right),$$

   where $\hat{\beta}_0$ and $\hat{\beta}$ are least squares estimates, which minimize

$$f(\beta_0, \beta) = \sum_{i=1}^{n}\left[Y_i - (\beta_0 + \beta_1 X_i)\right]^2.$$

   The first-order partial derivatives of $f(\beta_0, \beta)$ are

$$\begin{aligned} \frac{\partial f(\beta_0, \beta)}{\partial \beta_0} &= -2\sum_{i=1}^{n}\left[Y_i - (\beta_0 + \beta_1 X_i)\right]; \\ \frac{\partial f(\beta_0, \beta)}{\partial \beta} &= -2\sum_{i=1}^{n} X_i\left[Y_i - (\beta_0 + \beta_1 X_i)\right]. \end{aligned}$$

$\hat{\beta}_0$ and $\hat{\beta}$ should satisfy that

$$\begin{aligned}
\frac{\partial f(\hat{\beta}_0, \hat{\beta})}{\partial \beta_0} &= -2 \sum_{i=1}^{n} \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i \right) \right] = 0, \\
&\Rightarrow \sum_{i=1}^{n} \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i \right) \right] = \sum_{i=1}^{n} \hat{\epsilon}_i = 0. \\
\frac{\partial f(\hat{\beta}_0, \hat{\beta})}{\partial \beta} &= -2 \sum_{i=1}^{n} X_i \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i \right) \right], \\
&\Rightarrow \sum_{i=1}^{n} X_i \left[Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_i \right) \right] = \sum_{i=1}^{n} X_i \hat{\epsilon}_i = 0.
\end{aligned}$$

Hence,

$$\sum_{i=1}^{n} \hat{\epsilon}_i (\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n} \hat{\epsilon}_i \left(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{Y} \right) = \left(\hat{\beta}_0 - \bar{Y} \right) \sum_{i=1}^{n} \hat{\epsilon}_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i \hat{\epsilon}_i = 0.$$

Therefore,

$$TSS = ESS + RSS.$$

2. Show that under the assumption of normality, the likelihood ratio test for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ has the same form as the Wald test.

   Proof:

   In this problem, we have to assume that σ^2 is known, or the MLE $\hat{\sigma}^2$ is the same with the H_0 restriction or not. Here, we consider σ^2 as known.

   We know the MLEs of β_0 and β_1 as $\hat{\beta}_0$ and $\hat{\beta}_1$.

   Under H_0, the likelihood becomes

$$L(\beta_0, 0) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \beta_0)^2 \right\}.$$

   It can be easily derived that the restricted MLE of β_0 is $\bar{Y}$.

The likelihood ratio statistic is

$$\begin{aligned}
T = \frac{L(\hat{\beta}_0, \hat{\beta}_1)}{L(\bar{Y}, 0)} &= \frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \bar{Y})^2\right\}} \\
&= \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^n \left[(Y_i - \bar{Y})^2 - (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2\right]\right\} \\
(\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}) &= \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^n \left[(Y_i - \bar{Y})^2 - (Y_i - \bar{Y} + \hat{\beta}_1 \bar{X} - \hat{\beta}_1 X_i)^2\right]\right\} \\
&= \exp\left\{\frac{1}{2\sigma^2}\sum_{i=1}^n \left[2\hat{\beta}_1 (Y_i - \bar{Y})(X_i - \bar{X}) - \hat{\beta}_1^2 (X_i - \bar{X})^2\right]\right\} \\
\left(\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) &= \exp\left\{\frac{1}{2\sigma^2}\left[2\hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2 - \hat{\beta}_1^2 \sum_{i=1}^n (X_i - \bar{X})^2\right]\right\} \\
&= \exp\left\{\frac{\hat{\beta}_1^2}{2\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2\right\}
\end{aligned}$$

Note that

$$\lambda = 2\log T = \frac{\hat{\beta}_1^2}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2$$

has an approximately limiting distribution χ_1^2.

The Wald test statistic is

$$W = \frac{\hat{\beta}_1 - 0}{\text{se}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}{\sigma},$$

and has a limiting distribution $N(0, 1)$.

Since $Z^2 \sim \chi_1^2$ if $Z \sim N(0, 1)$, and $\lambda = W^2$, the likelihood ratio test is in the same form as the Wald test.

3. Consider the **regression through the origin** model:

$$Y_i = \beta X_i + \epsilon_i$$

(a) Find the least squares estimate for β.

(b) Find the standard error of the estimate.

(c) Find conditions that guarantee that the estimator is consistent.

Solution:

(a) To find the least squares estimate for β, we want to minimize

$$f(\beta) = \sum_{i=1}^{n} (Y_i - \beta X_i)^2.$$

By taking the first derivative and setting it as zero,

$$\frac{df(\beta)}{d\beta} = -2 \sum_{i=1}^{n} X_i (Y_i - \beta X_i) = 0$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2},$$

which is the least square estimate for β.

(b) The variance of $\hat{\beta}$ is

$$V[\hat{\beta}] = V\left[\frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}\right] = \frac{\sum_{i=1}^{n} X_i^2 V[Y_i]}{\left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{\sum_{i=1}^{n} X_i^2 V[\epsilon_i]}{\left(\sum_{i=1}^{n} X_i^2\right)^2}.$$

Assume that ϵ_i are i.i.d. with mean 0 and variance σ^2. Then

$$V[\hat{\beta}] = \frac{\sum_{i=1}^{n} X_i^2 \sigma^2}{\left(\sum_{i=1}^{n} X_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}.$$

(c) $\hat{\beta}$ is a consistent estimator of β iff

$$\Pr\left(|\hat{\beta} - \beta| > \delta\right) \to 0, \quad \forall \delta > 0.$$

Since

$$\hat{\beta} - \beta = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} - \beta = \frac{\sum_{i=1}^{n} X_i (\beta X_i + \epsilon_i)}{\sum_{i=1}^{n} X_i^2} - \beta = \frac{\sum_{i=1}^{n} X_i \epsilon_i}{\sum_{i=1}^{n} X_i^2}.$$

By Chebyshev's inequality,

$$\Pr\left(|\hat{\beta} - \beta| > \delta\right) = \Pr\left(\left|\frac{\sum_{i=1}^{n} X_i \epsilon_i}{\sum_{i=1}^{n} X_i^2}\right| > \delta\right) \leq \frac{V\left[\frac{\sum_{i=1}^{n} X_i \epsilon_i}{\sum_{i=1}^{n} X_i^2}\right]}{\delta^2} = \frac{\sigma^2}{\delta^2 \sum_{i=1}^{n} X_i^2}.$$

Hence, if $\frac{\sigma^2}{\delta^2 \sum_{i=1}^n X_i^2} \overset{n\to\infty}{\to} 0$, then $\Pr\left(|\hat{\beta} - \beta| > \delta\right) \to 0$.
That is, if there are infinite number of non-zero X_i's as $n \to \infty$, ($\sum_{i=1}^n X_i^2 \to \infty$), $\hat{\beta}$ is a consistent estimator.

4. Read in the data file `cars.dat` from bCourse.

   (a) Make a scatterplot of HP (horsepower) against MPG (miles per gallon); that is, with HP on the y-axis and MPG on the x-axis. Experiment with taking logs of one or both variables until you find a combination that looks appropriate for the simple linear regression model. Turn in your plot, along with an explanation of how you evaluated the assumptions of the model.

   (b) Using the transformations you chose in (a), fit a simple linear regression model. Report $\hat{\beta}_0$ and $\hat{\beta}_1$, and carry out a Wald test of size $\alpha = 0.05$ for $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$.

   (c) Make two diagnostic plots as follows, and turn in each one with a one sentence interpretation.

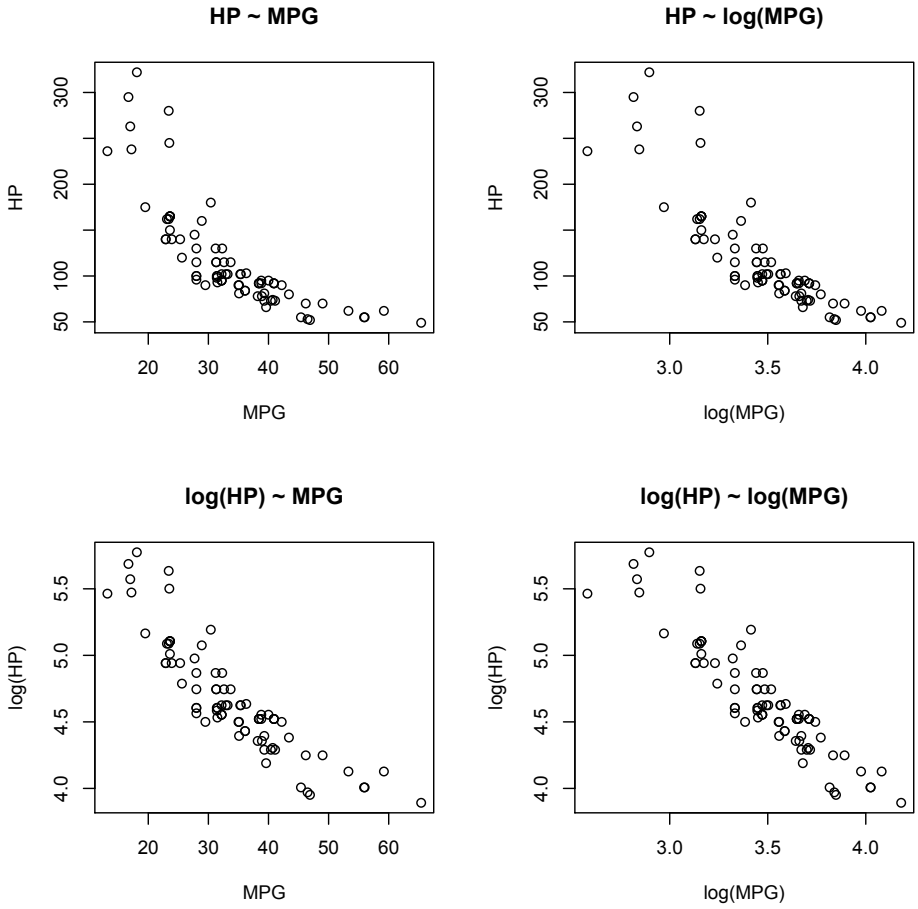
   - Plot the x values (for whatever transformation you used) on the x-axis, and the residuals on the y-axis. Do you see evidence against the assumption of constant variance?
   - Make a plot comparing the quantiles of the residuals to the quantiles of a normal distribution. (See the help for `R` functions `qqnorm` and `qqline`.) Do you see evidence against the assumption of normality?

Solution:

(a) We can make the scatterplots by using the following commands.

```
cars <- read.table("cars.dat", header=TRUE)
par(mfrow=c(2,2))
plot(x=cars$MPG, y=cars$HP, xlab="MPG", ylab="HP", main="HP ~ MPG")
plot(x=log(cars$MPG), y=cars$HP, xlab="log(MPG)", ylab="HP",
    main="HP ~ log(MPG)")
plot(x=cars$MPG, y=log(cars$HP), xlab="MPG", ylab="log(HP)",
    main="log(HP) ~ MPG")
plot(x=log(cars$MPG), y=log(cars$HP), xlab="log(MPG)", ylab="log(HP)",
    main="log(HP) ~ log(MPG)")
```

The four scatterplots are below.



From the scatterplots, we can see that log(HP) $\sim$ log(MPG) has the best linear pattern. Therefore, it is most appropriate to fit log(HP) $\sim$ log(MPG) with a simple linear model. That is

$$\log(\mathrm{HP}_i) = \beta_0 + \beta_1 \log(\mathrm{MPG}) + \epsilon_i,$$

$i = 1, \ldots, n$ and ϵ_i are i.i.d.s.

(b) We can fit the simple linear model by using the following `R` commands:

```
> linmod <- lm(log(HP)~log(MPG), data = cars)
> linmod$coefficients
(Intercept)    log(MPG)
   9.018889   -1.251056
```

So the estimates are $\hat{\beta}_0 \approx 9.02$ and $\hat{\beta}_1 \approx -1.25$.

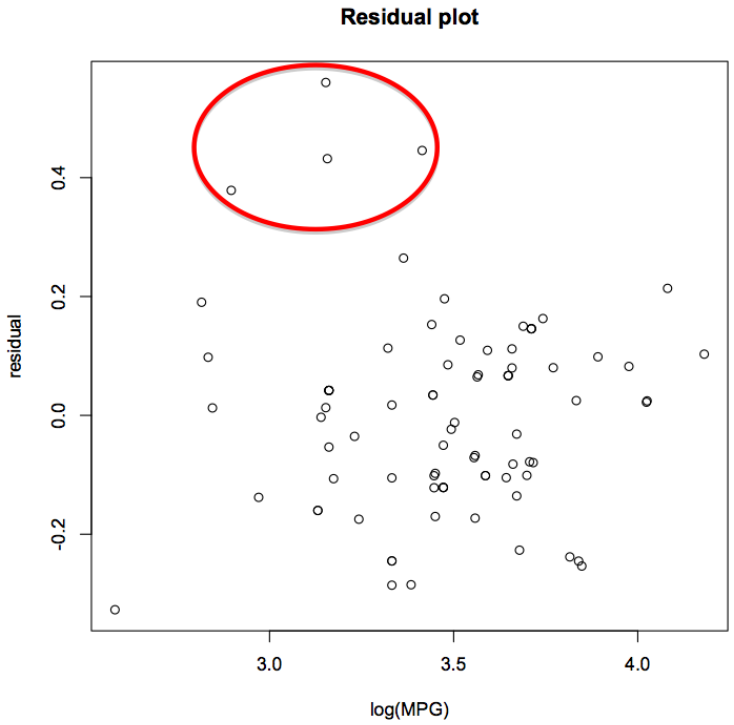
To carry out a Wald test of size $\alpha = 0.05$ for $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$, we can use the following `R` commands:

```
> beta1 <- linmod$coefficients[2]
> se.beta1 <- summary(linmod)$coefficients[2,2]
> W <- beta1/se.beta1
> 2*pnorm(-abs(W))
   log(MPG)
4.29821e-91
```

The p-value is extremely small, so we have a strong evidence to reject $H_0 : \beta_1 = 0$.

(c) The two diagnostic plots can be made as below.

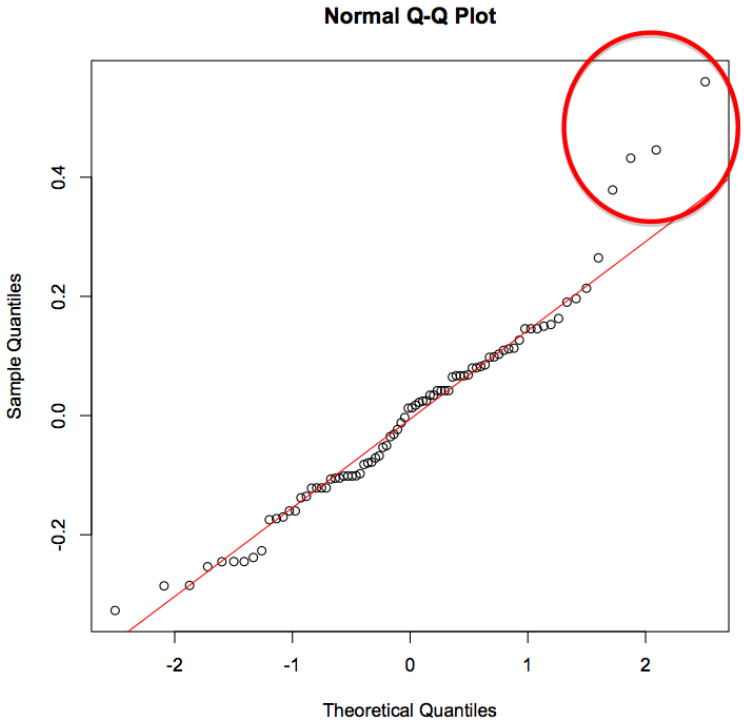
- `plot(x=log(cars$MPG), y=linmod$residuals, xlab="log(MPG)", ylab="residual", main="Residual plot")`



From the residual plot above, we can see that except for the 4 residuals which are possibly outliers in the read circle, the other residuals are

randomly scattered between $[-0.2, 0.2]$. This shows that assumption that ϵ_i's have mean zero and constant variance is reasonable.

- `qqnorm(linmod$residuals)`
  `qqline(linmod$residuals, col=2)`



The Q-Q norm plot shows that except for the same four outliers as in the residual plot, the rest of the residuals are approximately normally distributed. This shows that if we have a normality assumption on ϵ_i's, that assumption is not violated.

Please note that determining formally whether the four points are evidence against normality would require some idea of the variability in the sample quantiles under the null hypothesis of normality, which we have not discussed.

5. In this question we take a closer look at prediction intervals. Let $\theta = \beta_0 + \beta_1 X_*$, and let $\hat{\theta} = \hat{\beta}_0 + \hat{\beta}_1 X_*$. Thus, $\hat{Y}_* = \hat{\theta}$ while $Y_* = \theta + \epsilon$. Now, $\hat{\theta} \approx N(\theta, se^2)$, where

$$se^2 = V(\hat{\theta}) = V(\hat{\beta}_0 + \hat{\beta}_1 x_*).$$

Note that $V(\hat{\theta})$ is the same as $V(\hat{Y}_*)$. Now, $\hat{\theta} \pm 2\sqrt{V(\hat{\theta})}$ is an appropriate 95 percent confidence interval for $\theta = \beta_0 + \beta_1 x_*$ using the usual argument for a confidence interval. But, as you shall now show, it is not a valid confidence interval for Y_*.

(a) Let $s = \sqrt{V(\hat{Y}_*)}$. Show that

$$P(\hat{Y}_* - 2s < Y_* < \hat{Y}_* + 2s) \approx P\left(-2 < N\left(0, 1 + \frac{\sigma^2}{s^2}\right) < 2\right) \neq 0.95$$

(b) The problem is that the quantity of interest Y_* is equal to a parameter θ plus a random variable. We can fix this by defining

$$\xi_n^2 = V(\hat{Y}_*) + \sigma^2 = \left[\frac{\sum_i (x_i - x_*)^2}{n \sum_i (x_i - \bar{x})^2} + 1\right] \sigma^2$$

In practice, we substitute $\hat{\sigma}^2$ for σ^2 and we denote the resulting quantity by $\hat{\xi}_n$. Now consider the interval $\hat{Y}_* \pm 2\hat{\xi}_n$. Show that

$$P(\hat{Y}_* - 2\hat{\xi}_n < Y_* < \hat{Y}_* + 2\hat{\xi}_n) \approx P(-2 < N(0,1) < 2) \approx 0.95.$$

Solution:

(a) With the assumption that $\epsilon \sim N(0, \sigma^2)$, we have

$$Y_* \sim N(\theta, \sigma^2).$$

Also, based on the model

$$\hat{\beta}_0 \sim N(\beta_0, V[\hat{\beta}_0]); \quad \hat{\beta}_1 \sim N(\beta_1, V[\hat{\beta}_1]).$$

So

$$\hat{Y}_* = \hat{\beta}_0 + \hat{\beta}_1 X_* \sim N(\beta_0 + \beta_1 X_*, V[\hat{Y}_*]) = N(\theta, s^2).$$

Since $\hat{Y}_*$ is a linear combination of $\epsilon_1, \ldots, \epsilon_n$, and ϵ associated with the new observation Y_* is independent of $\epsilon_1, \ldots, \epsilon_n$, we have that Y_* and $\hat{Y}_*$ are independent. So

$$V[Y_* - \hat{Y}_*] = V[Y_*] + V[\hat{Y}_*] = \sigma^2 + s^2.$$

Hence,

$$\frac{Y_* - \hat{Y}_*}{s} \sim N\left(0, 1 + \frac{\sigma^2}{s^2}\right).$$

Therefore,

$$P(\hat{Y}_* - 2s < Y_* < \hat{Y}_* + 2s) = P\left(-2 < \frac{Y_* - \hat{Y}_*}{s} < 2\right)$$
$$\approx P\left(-2 < N\left(0, 1 + \frac{\sigma^2}{s^2}\right) < 2\right) \neq 0.95.$$

(b) Since

$$V(\hat{Y}_*) = V(\hat{\beta}_0 + \hat{\beta}_1 X_*) = V(\hat{\beta}_0) + X_*^2 V(\hat{\beta}_1) + 2X_* Cov(\hat{\beta}_0, \hat{\beta}_1)$$
$$= \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n\sum_{i=1}^n (X_i - \bar{X})^2} + X_*^2 \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} - 2X_* \frac{\sigma^2 \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2}$$
$$= \frac{\sigma^2 \sum_{i=1}^n (X_i - X_*)^2}{n\sum_{i=1}^n (X_i - \bar{X})^2},$$

$$V[Y_* - \hat{Y}_*] = V[Y_*] + V[\hat{Y}_*] = \sigma^2 + \frac{\sigma^2 \sum_{i=1}^n (X_i - X_*)^2}{n\sum_{i=1}^n (X_i - \bar{X})^2} = \left[\frac{\sum_i (X_i - X_*)^2}{n\sum_i (X_i - \bar{X})^2} + 1\right]\sigma^2 = \xi_n^2.$$

Hence,

$$\frac{Y_* - \hat{Y}_*}{\hat{\xi}_n} \overset{approx}{\sim} N(0, 1).$$

Therefore,

$$P(\hat{Y}_* - 2\hat{\xi}_n < Y_* < \hat{Y}_* + 2\hat{\xi}_n) = P\left(-2 < \frac{Y_* - \hat{Y}_*}{\hat{\xi}_n} < 2\right)$$
$$\approx P(-2 < N(0, 1) < 2) \approx 0.95.$$