2021F_Week06

September 27, 2021

0.1 Week06

Goals: - Write down the variance for $\hat{\beta}$ - Write down the distribution for $\hat{\beta}$ - Use the $\hat{beta}$ distribution to compute confidence intervals

0.1.1 Variance, $\mathcal{I}^{-1}$, for $\hat{\beta}_0$ and $\hat{\beta}_1$

We found that given a dataset of N pairs of points (x, y) the MLE for $\hat{\beta}_1$ was

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{1}$$

To compute the variance for _{1} we need to consider repeated sampels of our datset from random variables (X_i, Y_i). If we shift our focus from a fixed dataset to random variables that can generate our fixed data then the MLE is a funciton of random variables.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \tag{2}$$

Above we use capital X and capital Y to represent random variables. To be clear, $\text{Var}(y_i) = 0$ and $\text{Var}(Y_i) \geq 0$.

Variance of $\hat{\beta}_1$ Before we compute the variance of $\hat{\beta}_1$ we can simplify the numerator of the MLE

$$\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y}) \tag{3}$$

$$\sum_{i=1}^{N}(X_i - \bar{X})Y_i - (X_i - \bar{X})\bar{Y} \tag{4}$$

$$\sum_{i=1}^{N}(X_i - \bar{X})Y_i - \sum_{i=1}^{N}(X_i - \bar{X})\bar{Y} \tag{5}$$

$$\sum_{i=1}^{N}(X_i - \bar{X})Y_i - \bar{Y}\sum_{i=1}^{N}(X_i - \bar{X}) \qquad \bar{Y} \text{ is a constant} \tag{6}$$

$$\sum_{i=1}^{N}(X_i - \bar{X})Y_i \qquad \text{the second sum is zero} \tag{7}$$

Our MLE for $\hat{\beta}_1$ is now simpler

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})Y_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \tag{8}$$

We can think of the above as

$$\hat{\beta}_1 = \sum_{i=1}^{N} k_i Y_i \tag{9}$$

where we now consider

$$k_i = \frac{(X_i - \bar{X})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \tag{10}$$

a constant. That is, we assume we are given x_i values

$$k_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{11}$$

Our goal is to find the variance of $\hat{\beta}_1$ or

$$Var\left(\hat{\beta}_1\right) = Var\left(\sum_{i=1}^{N} k_i Y_i\right) \tag{12}$$

The variance of the sum (or differnece) of random variables We will need a fact about the variance of the sum of independent random variables.

For two independent random variables X and Y and constants a and b the following is true

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) \tag{13}$$

.

This also means that

$$Var(X - Y) = Var((1)X + (-1)Y) = 1^2 Var(X) + (-1)^2 Var(Y) = Var(X) + Var(Y) \tag{14}$$

.

The same relationship holds for N random variables $Y_1, Y_2, \cdots, Y_N$ which are all independent from one another and for constants $k_1, k_2, \cdots, k_N$

$$Var(a_1Y_1 + a_2Y_2 + \cdots + a_NY_N) = Var\left(\sum_{i=1}^{N} a_iY_i\right) = \sum_{i=1}^{N} a_i^2 Var\left(Y_i\right) \tag{15}$$

.

This mean then

$$Var\left(\hat{\beta}_1\right) = Var\left(\sum_{i=1}^{N} k_iY_i\right) \tag{16}$$

$$= \sum_{i=1}^{N} k_i^2 Var(Y_i) \tag{17}$$

However, our linear regression model states that every Y_i has the following same distribution

$$Y_i|x_i \sim \mathcal{N}\left(\beta_0 + \beta_1 x_i, \sigma^2\right) \tag{18}$$

That is, the variance for each Y_i is σ^2. Then the above equation simplifies to

$$Var\left(\hat{\beta}_1\right) = \sum_{i=1}^{N} k_i^2 \sigma^2 \tag{19}$$

$$= \sigma^2 \sum_{i=1}^{N} k_i^2 \tag{20}$$

$$\tag{21}$$

Our final step is to figure out what $\sum_{i=1}^{N} k_i^2$ equals.

$$k_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{22}$$

and if we square k_i we are left with

$$k_i^2 = \frac{(x_i - \bar{x})^2}{\left(\sum_{i=1}^{N}(x_i - \bar{x})^2\right)^2} \tag{23}$$

Summing up the k_i^2s we find

$$
\begin{aligned}
\sum_{i=1}^{N} k_i^2 &= \sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^2} && (24)\\
&= \frac{1}{\left(\sum_{i=1}^{N}(x_i-\bar{x})^2\right)^2}\sum_{i=1}^{N}(x_i-\bar{x})^2 && (25)\\
&= \frac{1}{\sum_{i=1}^{N}(x_i-\bar{x})^2} && (26)
\end{aligned}
$$

Variance of $\hat{\beta}_1$ The final result is the variance for $\hat{\beta}_1$

$$
Var(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \tag{27}
$$

Variance of $\hat{\beta}_0$ To compute the variance for $\hat{b}eta_0$ we can proceed the same way as we did for $\hat{\beta}_1$: (i) consider the random variables that generated the data, (ii) take the variance of the MLE (now a random variable), and (iii) simplify.

The MLE for $\hat{\beta}_0$ is

$$
\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x} \tag{28}
$$

The variance is then

$$
\begin{aligned}
Var(\hat{\beta}_0) &= Var(\bar{Y}) + \bar{x}^2 Var(\hat{\beta}_1) && \bar{x} \text{ is a constant} && (29)\\
&= Var\left(\frac{\sum_{i=1}^{N} Y_i}{N}\right) + \bar{x}^2 Var(\hat{\beta}_1) && \text{Definition of sample average} && (30)\\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var(Y_i) + \bar{x}^2 Var(\hat{\beta}_1) && \frac{1}{N} \text{is a constant} && (31)\\
&= \frac{1}{N^2}\sum_{i=1}^{N} \sigma^2 + \bar{x}^2 Var(\hat{\beta}_1) && \text{The variance of every } Y_i \text{ is } \sigma^2 && (32)\\
&= \frac{1}{N^2} N\sigma^2 + \bar{x}^2 Var(\hat{\beta}_1) && && (33)\\
&= \frac{1}{N^2} N\sigma^2 + \bar{x}^2\sigma^2 \frac{1}{\sum_{i=1}^{N}(x_i-\bar{x})^2} && && (34)\\
&= \sigma^2\frac{1}{N} + \sigma^2\frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2} && && (35)\\
&= \sigma^2\left(\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}\right) && && (36)\\
& && && (37)
\end{aligned}
$$

0.1.2 The distribution for $\hat{\beta}$

We know that as our dataset grows that the MLE for $\hat{\beta}_0$ and for $\hat{\beta}_1$ approaches a normal distribution centered at our true parameter values with mean parameter equal to the correpsonding MLEs and variance parameter equal to the corresponding variances we found above.

$$\hat{\beta}_0 \sim \mathcal{N}\left(\bar{Y} - \hat{\beta}_1\bar{x}, \sigma^2\left(\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right)\right) \tag{38}$$

$$\hat{\beta}_0 \sim \mathcal{N}\left(\sum_{i=1}^{N}\frac{(x_i - \bar{x})y_i}{(x_i - \bar{x})^2}, \sigma^2\frac{1}{\sum_{i=1}^{N}(x_i - \bar{x})^2}\right) \tag{39}$$

0.1.3 Computing confidence intervals for $\hat{\beta}$

The equation to compute a $1 - \alpha$ confidence interval for a normally distributed random variable with mean equal to μ and variance equal to σ^2 is

$$\text{CI}_{1-\alpha} = (\mu - z_{1-\alpha/2}\sigma, \mu + z_{\alpha/2}\sigma) \tag{40}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile value from a random variable $Z \sim \mathcal{N}(\imath, \infty)$. The p quantile value q is the value q such that $P(Z < p) = q$

α	$z_{1-\alpha/2}$
0.01	2.58
0.05	1.96
0.10	1.64
0.20	1.28

0.1.4 Why this formula works (Shifting and scaling Normal distributions)

For a random variable $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ We can derive the distribution of a second random variable $Y = X + c$ where c is a constant. Adding or subtracting a constant from a normally distributed random variable,X,—shifting—creates a new random variable Y that is normally distributed

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \tag{41}$$
$$\text{if } Y = X + c \tag{42}$$
$$\text{then } Y \sim \mathcal{N}\left(\mu + c, \sigma^2\right) \tag{43}$$

Multiplying a random variable X by a constant c—scaling—creates a new random variable Y that is normally distributed

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right) \tag{44}$$
$$\text{if } Y = cX \tag{45}$$
$$\text{then } Y \sim \mathcal{N}\left(\mu, c^2\sigma^2\right) \tag{46}$$

0.1.5 From a $N(0,1)$ to any $N(\mu, \sigma^2)$

Let $Z \sim \mathcal{N}(0,1)$. Then we can transform Z into Y, a $\mathcal{N}\left(0, \sigma^2\right)$, by mutliplying by the constant σ

$$\text{if } Y = \sigma Z \text{ then } Y \sim \mathcal{N}\left(0, \sigma^2 \times 1\right) \tag{47}$$
$$Y \sim \mathcal{N}\left(0, \sigma^2\right) \tag{48}$$

Then we can create a normal distribution, X, centered at μ by shifting Y by the constant μ

$$\text{if } X = Y + \mu \text{ then } Y \sim \mathcal{N}\left(\mu, \sigma^2\right) \tag{49}$$

We transformed Z, a $\mathcal{N}(0,1)$ random variable, into X, a $\mathcal{N}(\mu, \sigma^2)$ random variable by first multiplying Z by σ and then adding μ

$$X = \mu + \sigma Z \tag{50}$$

This also means that any $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ can be converted to a $Z \sim \mathcal{N}(0,1)$ by running the above process in reverse

$$Z = \frac{X - \mu}{\sigma} \tag{51}$$

0.1.6 Confidence intervals

We will compute a $1 - \alpha$ confidence interval for a random variable X by computing two quantiles values: (i) the quantile value l such that $P(X < l) = \alpha/2$ and the quantile value u such that $P(X < u) = 1 - \alpha/2$.

We can leverage the above transformation to make computing these two quantiles easier **as long as our random variable is normally distributed**. The idea is to compute the quantile values for Z and transform those values so that they are confidence intervals for an arbitrary $\mathcal{N}(\mu, \sigma^2)$ random variable.

For the lower quantile value:

$$
\begin{align}
P(Z < l) &= \alpha/2 \tag{52} \\
P\left(\frac{X-\mu}{\sigma} < l\right) &= \alpha/2 \tag{53} \\
P(X - \mu < l\sigma) &= \alpha/2 \tag{54} \\
P(X < \mu + l\sigma) &= \alpha/2 \tag{55} \\
& \tag{56}
\end{align}
$$

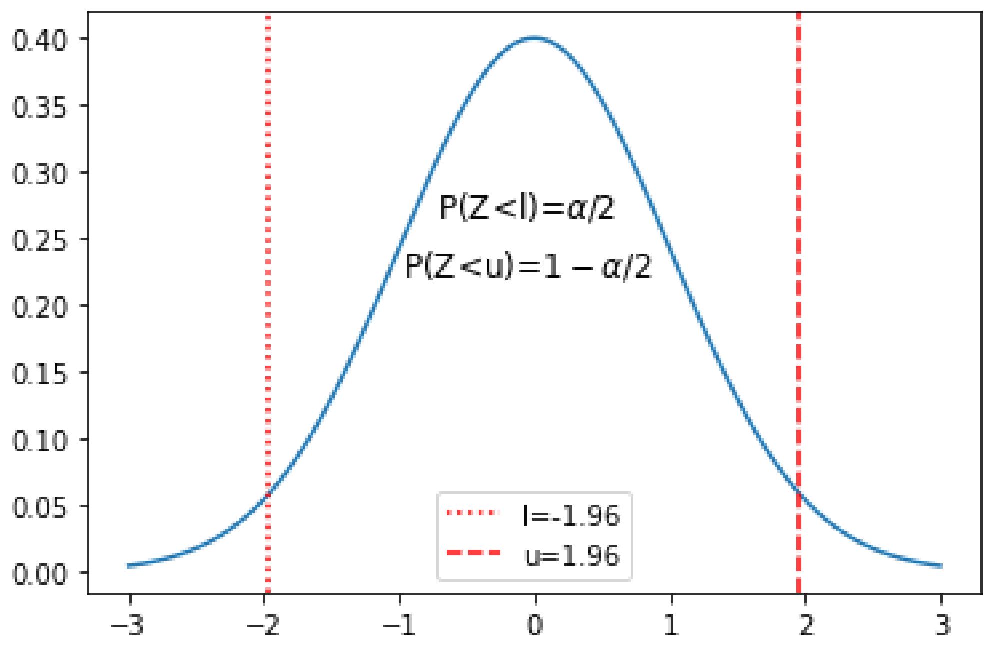
For the upper quantile value

$$
\begin{align}
P(Z < u) &= 1 - \alpha/2 \tag{57} \\
P\left(\frac{X-\mu}{\sigma} < u\right) &= 1 - \alpha/2 \tag{58} \\
P(X - \mu < u\sigma) &= 1 - \alpha/2 \tag{59} \\
P(X < \mu + u\sigma) &= 1 - \alpha/2 \tag{60} \\
& \tag{61}
\end{align}
$$

One final step is recognizing that because the Normal distribution is symmetric the quantile value l and u are the same value but opposite sign.

[22]:
```
import scipy.stats

dom = np.linspace(-3,3,10**3)

normal = scipy.stats.norm(0,1)

fig,ax = plt.subplots()
ax.plot( dom,normal.pdf(dom) )

alpha = 0.05
l = normal.ppf(alpha/2)
u = normal.ppf(1-alpha/2)

ax.axvline(l,color="red",linestyle=":",label="l=-1.96")
ax.axvline(u,color="red",linestyle="--",label="u=1.96")

ax.text( 0.50,0.70,r"P(Z<l)=$\alpha/2$  ",transform=ax.
 ↪transAxes,ha="center",va="top",fontsize=12  )
ax.text( 0.50,0.60,r"P(Z<u)=$1-\alpha/2$  ",transform=ax.
 ↪transAxes,ha="center",va="top",fontsize=12  )

ax.legend()
```

[22]:
```
<matplotlib.legend.Legend at 0x1b60eb310>
```



This means that for a normal distribution $u = -l$, and so

$$P(X < \mu + l\sigma) = \alpha/2 \tag{62}$$
$$P(X < \mu + u\sigma) = 1 - \alpha/2 \tag{63}$$
$$\tag{64}$$

becomes

$$P(X < \mu - u\sigma) = \alpha/2 \tag{65}$$
$$P(X < \mu + u\sigma) = 1 - \alpha/2 \tag{66}$$
$$\tag{67}$$

or

$$P(\mu - u\sigma < X < \mu + u\sigma) = 1 - \alpha/2 - \alpha/2 = 1 - \alpha \tag{68}$$
$$\tag{69}$$

0.1.7 Application to $\hat{\beta}$

Suppose we are given the following data set:

x	y
1	0
2	3
3	6
4	7
5	2

We can compute $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})y_i}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{70}$$

$$= \frac{\sum_{i=1}^{N}(x_i - 3)y_i}{\sum_{i=1}^{N}(x_i - 3)^2} \tag{71}$$

$$= 0.8 \tag{72}$$

and we can compute the variance of $\hat{\beta}_1$ by first finding$\hat{\sigma}^2$ and then computing the MLE of the variance:

$$Var(\hat{\beta}_1) = \bar{\sigma}^2 \frac{1}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{73}$$

From the above data you can find $\hat{\sigma}^2$ equals 6.7 and so

$$Var(\hat{\beta}_1) = 5.36 \frac{1}{\sum_{i=1}^{N}(x_i - 3)^2} \tag{74}$$

$$= 0.536 \tag{75}$$

$$\tag{76}$$

or

$$sd(\hat{\beta}_1) = \sqrt{0.536} = 0.732 \tag{77}$$

$$\tag{78}$$

Then our distribution for $\hat{\beta}_1$ is

$$\hat{\beta}_1 \sim \mathcal{N}(0.8, 0.732^2) \tag{79}$$

We can compute a 95% confidence interval (or 1-0.05 confidence interval) using this information and our table of quantile values above.

$CI(\hat{\beta}_1)_{1-0.05} = (0.8 - 1.96 * 0.732, 0.8 + 1.96 * 0.732)$