Math 455 Class 11 Feb. 13

Understanding regression coefficients.

Change in Pockets: sample of n people

everyone has 0 – 4 quarters

0 – 4 dimes

0 – 4 nickels

0 – 4 pennies

For a total Cents of 0 – 164.

We define total Coins = # of coins: 0–16

small coins = # pennies + # nickels

Regression: total Cents ~ total Coins + small coins.

$$\text{total } \cancel{\text{Coins}}\ \text{Cents} = \beta_0 + \beta_1 \text{ totalCoins} + \beta_2 \text{ smallCoins} + \varepsilon$$

What values do you expect ~~from~~ for $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$?

Now what if we do:

$$\text{total Cents} = \cancel{\beta_0 +} \ \alpha_0 + \alpha_1 \text{ small Coins} + \varepsilon$$

This is because ~~reducing small~~ increasing small Coins while holding total Coins constant means trading a quarter/dime for a nickel/penny reducing total Cents.

We looked at this with R.

To explain the results, remember that

- $(\bar{Y}, \bar{X}_1, \bar{X}_2)$ is on the regression plane:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2.$$

- The residuals sum to 0, or:

$(Y - \hat{Y})$ is perp to the ~~pla~~ sub-space of R^n generated by $\vec{1}, X_1, X_2$.

Q: In the 1-variable case, the p-value is connected to the F-statistic, which is connected to R^2.

Is it possible to have statistical significance and a low R^2 at the same time?

A: YES. See "Small Coins" model.

There is a difference between "statistical significance" and "practical significance".

Proof of Gauss–Markov Explained (Faraway p. 22)

ψ: an estimable function of parameters $\beta_0, \ldots, \beta_{p-1}$

$$\psi = c^T\beta = c_0\beta_0 + \cdots + c_{p-1}\beta_{p-1}$$

Think of an example: $c = (0, 1, 0, \ldots, 0)^T$

$$\psi = 0\beta_0 + 1\beta_1 + \cdots + 0\beta_{p-1} = \beta_1.$$

$\hat{\psi} = c^T\hat{\beta}$ the OLS estimator of ψ.

In the example, $\hat{\psi} = 0\hat{\beta}_0 + 1\hat{\beta}_1 + \cdots + 0\hat{\beta}_{p-1} = \hat{\beta}_1$.

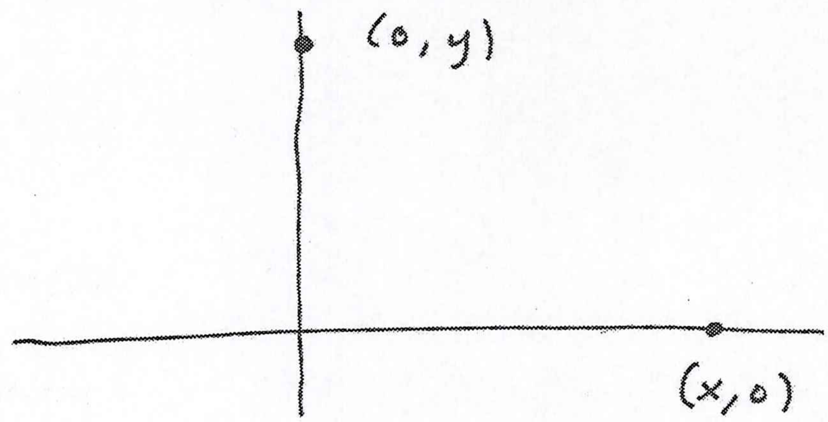
Let a^Ty be <u>any</u> unbiased linear estimator of ψ.

Gauss-Markov says "$\hat{\psi}$ is best" which means

that $V[a^Ty] \geq V[\hat{\psi}]$.

Remember: B in BLUE means <u>Best</u>

Best means "minimum variance" among unbiased linear estimators of ψ.



$y \geq 0$

What point on the x-axis is closest to $(0, y)$? (With proof.)

Answer: the origin $(0, 0)$.

Why? the distance from $(0, y)$ to an arbitrary point $(x, 0)$ on the x-axis is

$$d = \sqrt{(0-x)^2 + (y-0)^2} = \sqrt{x^2 + y^2}$$

Minimizing d is the same as minimizing d^2 (You'll get the same point.)

$$d^2 = y^2 + x^2 = d^2((0,y),(0,0)) + x^2$$

$$\geq d^2((0,y),(0,0))$$

since $x^2 \geq 0$.

$$E[a^T Y] = a^T E[Y] = a^T E[X\beta + \varepsilon]$$

$$= a^T\left(E[X\beta] + \cancel{E[\varepsilon]}^{\,0}\right) = a^T E[X\beta]$$

$$= a^T X\beta.$$

$E[a^T Y] = \psi = c^T\beta$ by unbiasedness.

Thus for any β, $a^T X\beta = c^T\beta$.

Thus $a^T X = c^T$ by linear algebra.

<u>Step 2</u>: If we assume $X^T X$ is invertible, this is immediate. But the result holds even without this assumption, because by step 1, $X^T a = c$ and range $(X^T X) =$ range (X^T) by linear algebra, so $c \in$ range $(X^T X)$.

<u>Step 3</u>: $c^T\hat{\beta} = \lambda^T X^T X\hat{\beta} = \lambda^T (X^T X)(X^T X)^{-1} X^T Y$

$$= \lambda^T X^T Y.$$

[Recall $\hat{\beta} = (X^T X)^{-1} X^T Y$]

Plan to show that $V[a^T Y] \geq V[\hat{\psi}]$:

Write $V[a^T Y] = V[\hat{\psi}] + V[\text{something}]$.

Variance is ≥ 0, so the result follows.

<u>Step 1</u>: $a^T X = c^T$

<u>Step 2</u>: $\exists \lambda$ so that $c = X^T X \lambda$.

<u>Step 3</u>: $c^T \hat{\beta} = \lambda^T X^T Y$.

<u>Step 4</u>: $V[a^T Y] = V[a^T Y - c^T \hat{\beta}] + V[c^T \hat{\beta}]$

Since $\hat{\psi} = c^T \hat{\beta}$, step 4 proves the result.

Why do steps 1-4 work?

We examine these individually.

Step 1 is true because we assumed that $a^T Y$ is an unbiased estimator of ψ.

Step 4: It is enough to show that

$$\operatorname{cov}\left[a^T Y - c^T \hat{\beta},\ c^T \hat{\beta}\right] = 0.$$

By step 3, we can replace $c^T \hat{\beta}$ with $\lambda^T X^T Y$.

$$\operatorname{Cov}\left[a^T Y - \lambda^T X^T Y,\ \lambda^T X^T Y\right]$$

$$= \operatorname{Cov}\left[(a^T - \lambda^T X^T) Y,\ \lambda^T X^T Y\right]$$

$$= (a^T - \lambda^T X^T) \operatorname{Cov}[Y, Y]\, X\lambda$$

$$= (a^T - \lambda^T X^T) \operatorname{Cov}\left[X\beta + \varepsilon,\ X\beta + \varepsilon\right] X\lambda.$$

$$= (a^T - \lambda^T X^T) \operatorname{Cov}[\varepsilon, \varepsilon]\, X\lambda$$

$$= (a^T - \lambda^T X^T)\ \sigma^2 I\ X\lambda.$$ by hyp. of lin. regression.

$$= (a^T X - \lambda^T X^T X)\ \sigma^2 I \lambda$$

$$= (c^T - c^T)\ \sigma^2 I \lambda$$ by Step 1 & 2.

$$= 0.$$