Appendix

1. Omitted Variables Bias

True model: $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

$E(\varepsilon | X_1, X_2) = 0$

$X_2 = X_1\delta_1 + u$

Omit X_2: $Y = X_1\widetilde{\beta_1} + \widetilde{\varepsilon}$

$E(\widetilde{\beta_1}) = \beta_1 + \delta_1\beta_2$

$\therefore$

No problem:	$\beta_2 = 0$	or	$\delta_1 = 0$
Problem:	$\beta_2 \neq 0$	and	$\delta_1 \neq 0$

2. Measurement Error

(i) $Y = Y^* + e$ (Y^*: true value, e: ME, Y: observed value)

True model: $Y^* = X \cdot \beta + \varepsilon$, $E(\varepsilon|X) = 0$

With ME: $Y = X \cdot \tilde{\beta} + \tilde{\varepsilon}$

$Y = X \cdot \beta + \varepsilon + e$

If $E(e|X) = 0$. ① unbiased and consistent $\hat{\beta}$
(e与Y相关，与Y^*不相关) ② Inference problem: $Var(\varepsilon + e) > Var(\varepsilon)$
If $E(e|X) \neq 0$ ① biased and inconsistent $\hat{\beta}$
(e与Y^*相关，与Y不相关) ② 〃

(ii) $X = X^* + e$ (X^*: true value, e: ME, X: observed value)

True model: $Y = X^* \beta + \varepsilon$

With ME: $Y = X \tilde{\beta} + \tilde{\varepsilon}$

$$Y = (X - e)\beta + \varepsilon$$
$$= X\beta + (\varepsilon - e\beta)$$

If $E(e|X) = 0$ ① unbiased and consistent $\hat{\beta}$
(e与X不相关，与X^*相关) ② Inference problem
If $E(e|X) \neq 0$. [Classical Errors-in-Variables (CEV) problem]
(e与X^*不相关，与X相关) $\hat{\beta} = \frac{cov(Y,X)}{Var(X)} = \frac{cov(X\beta + \varepsilon - e\beta, X)}{Var(X)} = \beta\left(1 - \frac{cov(e,X)}{Var(X)}\right)$

$$= \beta\left(1 - \frac{\sigma_e^2}{\sigma_{x^*}^2 + \sigma_e^2}\right)$$

$\therefore |\hat{\beta}| < |\beta|$ attenuation bias

3. Power (in lab / field experiment)

power = 1 − type II error　该拒绝时拒绝的概率.

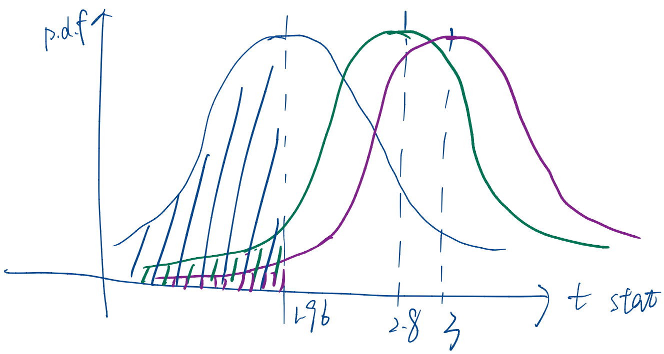
Data: $\{y_{11}, y_{12}, \cdots, y_{1i}, \cdots y_{1n}\}$, $\{y_{21}, \cdots, y_{2i}, \cdots, y_{2n}\}$, $Var(y_{ji}) = \sigma_y^2$

$$Var(\Delta\bar{y}) = Var(\bar{y}_1 - \bar{y}_2) = Var(\bar{y}_1) + Var(\bar{y}_2) = \frac{2}{n}\sigma_y^2$$

$$\therefore \hat{SD}(\Delta\bar{y}) = \sqrt{\frac{2}{n}}\,\sigma_y$$

$$\therefore E(t\ stat) = E\left(\frac{\Delta\bar{y}}{SD(\Delta\bar{y})}\right) = \sqrt{\frac{n}{2}}\cdot\frac{\Delta y}{\sigma_y} \xrightarrow{p} \text{Normal distr.}$$

- If $E(t\ stat) = 1.96$, then power $= Pr(t > 1.96) = 50\%$
- 〃 2.8, 〃 〃 $= 80\%$
- 〃 3, 〃 〃 $= 90\%$



∴ $\boxed{(n, \alpha, \beta)}$ 已知其二，可推余下的.

- n → sample size
- α → confidence level
- β → power

4. Regression table

$$MS = \frac{SS}{df}$$

. regress price mpg rep78

Source	SS	df	MS
Model	144754063	2	72377031.7
Residual	432042896	66	6546104.48
Total	576796959	68	8482308.22

$$\frac{\sum \hat{y}_i^2}{k-1}$$

$$\frac{\sum e_i^2}{n-(k-1)}$$

$$\frac{\sum y_i^2}{n-1}$$

Number of obs	=	69
F(2, 66)	=	11.06
Prob > F	=	0.0001
R-squared	=	0.2510
Adj R-squared	=	0.2283
Root MSE	=	2558.5

$$\sqrt{\frac{\sum e_i^2}{n-(k-1)}}$$

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
mpg	-271.6425	57.77115	-4.70	0.000	-386.9864	-156.2987
rep78	666.9568	342.3559	1.95	0.056	-16.5789	1350.492
_cons	9657.754	1346.54	7.17	0.000	6969.3	12346.21

$$\approx \left[\hat{\beta} - 1.96 \times SE(\hat{\beta}),\ \hat{\beta} + 1.96 \times SE(\hat{\beta})\right]$$