Valid Post-Selection Inference (Berk 2013)

① bound below with intersection
② improve critical value in t test.

Addressing problems: ① inference is not independent with model selection
② meaning of parameters change in every submodel. (# explaination $\sim p2^{p-1}$)

Setting: the true model is $Y \sim N(\mu, \sigma^2 I_p)$

inference target $\beta_M = \underset{\beta}{\arg\min} \|\mu - X_M \beta\|^2$ $\quad X_M$ is subset of $X_{n\times p}$ $\quad |M| = m$, $M \subset \{1, \dots, p\}$

Assume: ① β_M is not true model. $X_M \beta_M = \mu_M \neq \mu$ is approximation of μ.
First order correctness ~~?~~ doesn't hold.

② β_{M_F} is neither correct. Full model has no special status.

③ $\beta_{\setminus M} \neq 0$, $\beta_{\setminus M}$ doesn't exist (not needing to do inference on)

④ homoscedasticity: σ^2 constant for all sample.
there is method to estimate $\hat{\sigma}^2$ independently from model selection.
with ~~freedom r.~~ dof r. $\hat{\sigma}^2 \sim \chi_r^2$

Construction: $\beta_{j\cdot M} = \dfrac{X_{j\cdot M}^T \mu}{\|X_{j\cdot M}\|^2}$ $\quad \hat{\beta}_{j\cdot M} = \dfrac{X_{j\cdot M}^T Y}{\|X_{j\cdot M}\|^2} \sim N\left(\beta_{j\cdot M}, \dfrac{\sigma^2}{\|X_{j\cdot M}\|^2}\right)$

$t_{j\cdot M} = \dfrac{\hat{\beta}_{j\cdot M} - \beta_{j\cdot M}}{((X_M^T X_M)^{-1})_{jj}^{\frac{1}{2}} \hat{\sigma}} = \dfrac{(Y-\mu)^T X_{j\cdot M}}{\hat{\sigma}\|X_{j\cdot M}\|} \sim t_r$ $\qquad$ distribution of $t_{j\cdot \hat{M}}$ is distorted.

known model M / marginally $>$ $CI_{j\cdot M}(K) \triangleq [\hat{\beta}_{j\cdot M} \pm K[(X_M^T X_M)^{-1}]_{jj}^{\frac{1}{2}} \hat{\sigma}]$

target: $P(\forall j \in \hat{M}, \beta_{j\cdot \hat{M}} \in CI_{j\cdot \hat{M}}(K)) \geq 1-\alpha$ $\qquad$ randomness in $\hat{M}$ comes from ~~M~~ Y.

$\hat{M}(Y)$ only depends on distribution of Y, not sample.

sol: Find a K, s.t. ~~$\forall \hat{M}$~~ $P(\forall \hat{M}, \forall j \in \hat{M}, \beta_{j\cdot \hat{M}} \in CI_{j\cdot \hat{M}}(K)) \geq 1-\alpha$

K is the only thing needs to be specified. and was different for every $\hat{M}$.

Value of K: s.t. $P(\max\limits_{M \in \mathcal{M}} \max\limits_{j \in M} |t_{j\cdot M}| \leq K) \geq 1-\alpha$ $\qquad K = K(X, \mathcal{M}, \alpha, r)$.

$\forall \hat{M}$, $\max\limits_{j \in \hat{M}(Y)} |t_{j\cdot \hat{M}(Y)}(Y)| \leq \max\limits_{M \in \mathcal{M}} \max\limits_{j \in M} |t_{j\cdot M}(Y)| \Rightarrow$ ~~$P(\forall j \in \hat{M}, \beta_{j\cdot \hat{M}} \in$~~
$P(\max\limits_{j \in \hat{M}(Y)} |t_{j\cdot \hat{M}(Y)}(Y)| \leq K) \geq 1-\alpha$

Note: upper bound is sharp.
sharp model selection: $\hat{M}_{SPAR}(Y) \triangleq \underset{M \in \mathcal{M}}{\arg\max} \max\limits_{j \in M} |t_{j\cdot M}(Y)|$ "significance hunting".
① in selected model. "max j" significance was boosted by other less significant covariates.
② care nothing about model fit.

Compare with Scheffe's method. $\qquad d = \operatorname{rank}(X)$

$P(\sup\limits_{x \in \operatorname{span}(X)} |\dfrac{(Y-\mu)^T x}{\hat{\sigma}\|x\|}| \leq K_{Sch}) = 1-\alpha$ $\qquad K_{sch} = K_{sch}(\alpha, d, r) = \sqrt{d F_{d,r,1-\alpha}}$

$K(X, \mathcal{M}, \alpha, r) \leq K_{Sch}(\alpha, d, r) \underset{r\to+\infty}{\sim} \sqrt{d}$

if X orthogonal $\quad K_{orth} \sim \sqrt{2\log d}$ $\quad$ but if not, can be as bad as Scheffe (for worst case)