Data Analysis and Statistical Inference
Dr. Çetinkaya-Rundel
Duke University

In-Video Quiz Questions for Unit 7: Part 3 – (1) Model Selection

(05:20) – slide 7, after "Which variable, if any, should be dropped from the model first when doing backwards elimination using the p-value approach?"

1. The following model uses data from the American Community Survey to predict income from hours worked per week, race, and gender. Which variable (if any) should be dropped from the model first when doing backwards elimination using the p-value approach?

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2782.5726	6676.5534	0.42	0.6770
hrs_work	1247.2128	146.2013	8.53	0.0000
race:black	-9565.3090	6393.2168	-1.50	0.1350
race:asian	35816.6156	8690.3484	4.12	0.0000
race:other	-11112.8617	7213.3220	-1.54	0.1238
gender:female	-16430.0916	3803.4700	-4.32	0.0000

(a) race: black
(b) race: other
(c) race variable as a whole
(d) gender
(e) none

(09:54) – slide 9, after "We repeat until any of the remaining variables do not have a significant p-value."

2. If using the p-value forward selection approach, which variable should be added to the model first?

```
lm(formula = kid_score ~ mom_hs, data = cognitive)

            Estimate Std. Error t value Pr(>|t|)
mom_hsyes     11.771      2.322   5.069 5.96e-07
```

```
lm(formula = kid_score ~ mom_iq, data = cognitive)

            Estimate Std. Error t value Pr(>|t|)
mom_iq       0.60997    0.05852   10.42  < 2e-16
```

```
lm(formula = kid_score ~ mom_work, data = cognitive)

            Estimate Std. Error t value Pr(>|t|)
mom_workyes    5.832      2.552   2.285   0.0228
```

```
lm(formula = kid_score ~ mom_age, data = cognitive)

            Estimate Std. Error t value Pr(>|t|)
mom_age       0.6952     0.3620   1.920   0.0555
```

(a) mom_hs
(b) mom_iq
(c) mom_work
(d) mom_age

Answers:

1. e

Explanation: We can't drop individual levels of a variable, and since at least one level of race is significant, we keep the whole variable in.

2. b

Explanation: mom_iq has the lowest p-value.