DATE: 4/21/2022
TOPIC: test 2 con.

QUESTIONS:

IDEAS:

ACTIONS:

1) a, c

2) a, d

5) ⓒ for every drink per day, life expectancy decreases at a rate of 3.265 years.

ⓓ yes there is, because both predictors and the intercept have a Pr (>|t|) less than 0.05, and therefore all equal. this means the odds of any of the predictors and the response have less than a 5% chance of taking on their values entirely randomly. from there, we understand because the odds are equivalent across variables & there are no standouts, this means no one variable is more significant than the other.

6) ⓒ I would choose model 3 because the interaction term would account for any dependency / relationship between weight & transmission type, e.g. if newer automatic cars tend to be lighter in weight. additionally, model 3 meets the assumption of linearity, unlike 1 & 2. 1 & 2 both show some curvature on their residual plots, unlike 3.

ⓓ

$H_0: \beta_1 = \beta_2 = \beta_3 = 0$ p-value (model 3) = $1.669e^{-12}$

$H_a: \beta_1 \neq \beta_2 \neq \beta_3 \neq 0$ based on the p-value of $1.669e^{-12}$, we are 95% confident that β_3 is greater than 0, and is therefore statistically significant in the model. so, we reject the null hypothesis

$\alpha = 0.05$

↑ ⓐ $H_0: \beta_2 = \beta_3 = 0$

H_a: at least $\beta_2 \neq 0$ or $\beta_3 \neq 0$ or both

Transfer to your planner

DATE:

TOPIC:

QUESTIONS:

IDEAS:

ACTIONS:

- [ ]
- [ ]
- [ ]
- [ ]
- [ ]
- [ ]
- [ ]
- [ ]
- [ ]
- [ ]

Transfer to your planner!

(b) 6b provides a more reliable test because it is a test of model 3 only, while 7a is a nested F test of models 1 & 3, where model 1 is reduced & model 3 is full. however, the residual plot in model 1 violates the linearity assumption, so it would be best to just use model 3.

8) - after adding the variable, I would check the R^2 value & compare it to the model 3 R^2. if the R^2 was higher than it was without the additional variable, I would leave it in. this is to check the variable is actually significant to the test.
- then I would try backward elimination to identify which variable has the largest p-value, re-fitting the model until all the predictors are significant.
- then I would try eliminating predictors that give the largest drop in Mallow's C_p, until C_p doesn't decrease after predictor removal.