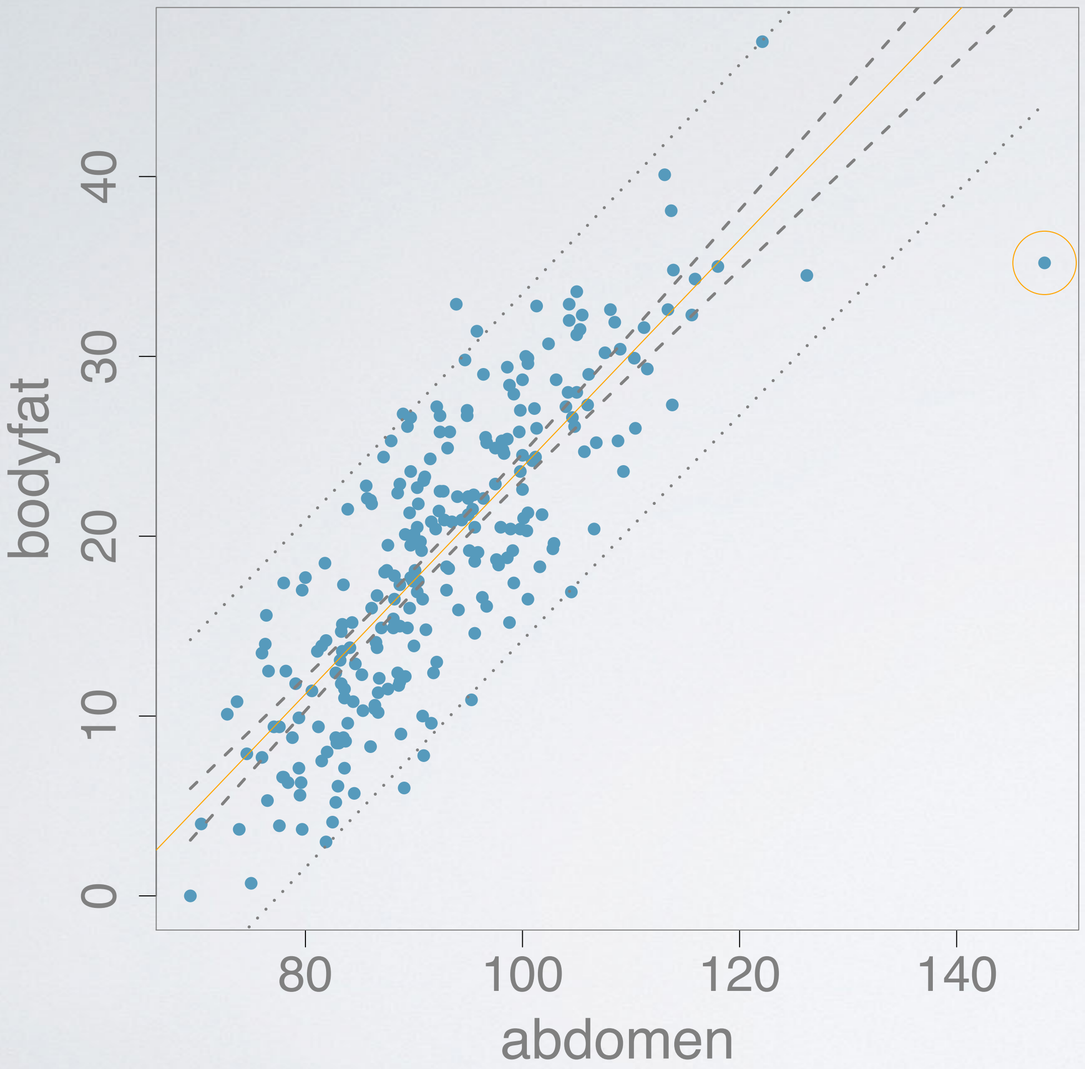
checking for outliers

Dr. Merlise Clyde

Duke
UNIVERSITY

prediction intervals



outliers?

$\varepsilon = y - (\alpha + \beta x)$

outlier probability: $\mathrm{P}(|\varepsilon| > k\sigma \mid \mathrm{data})$

- $k = 3$
- $\mathrm{P}(\mathrm{case\ 39\ is\ an\ outlier} \mid \mathrm{data}) = 0.9917$
- $k = 3.71$ prior probability any outliers is 0.05
- $\mathrm{P}(\mathrm{case\ 39\ is\ an\ outlier} \mid \mathrm{data}) = 0.6848$
- case 39 has a high probability of being from a different population

Source: https://pixabay.com/en/hulk-green-strong-big-muscle-male-1132281/

summary

- review from last video
- check for outliers
- rethink modeling assumptions