statistics in R

workshop 3

kim nguyen 3/22/2022

statistical tests

- categorical
  - goodness of fit for one categorical variable
  - t-tests
    - analysis of variance (anova)
- continuous
  - correlation and regression- linear relationships
- general linear models

types of data

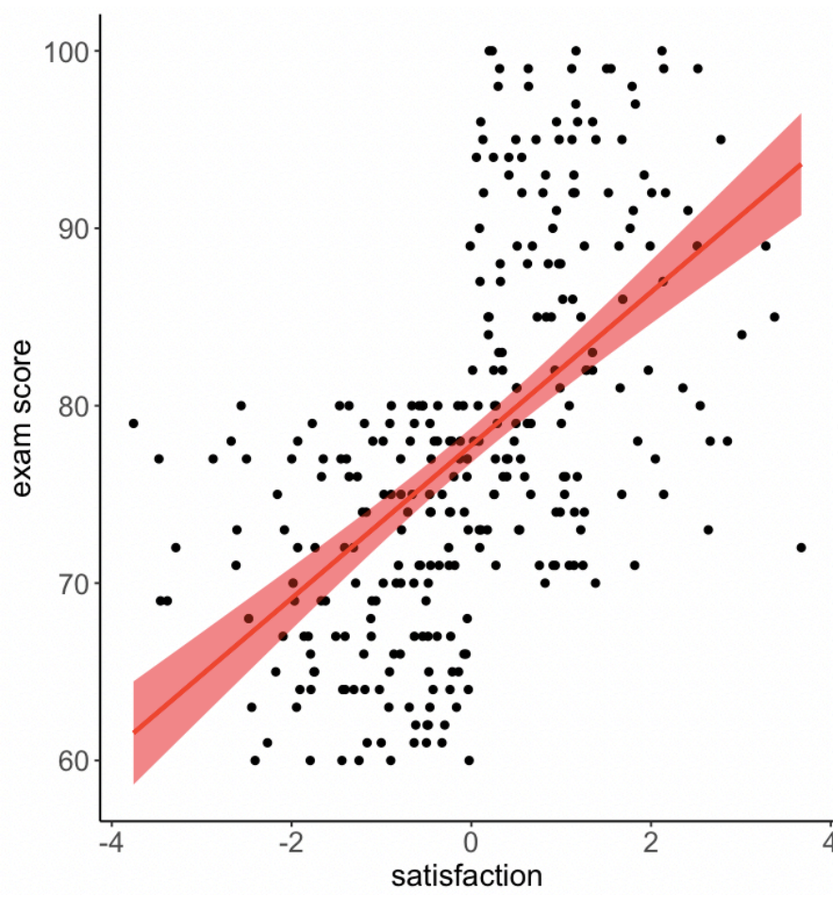
- continuous data

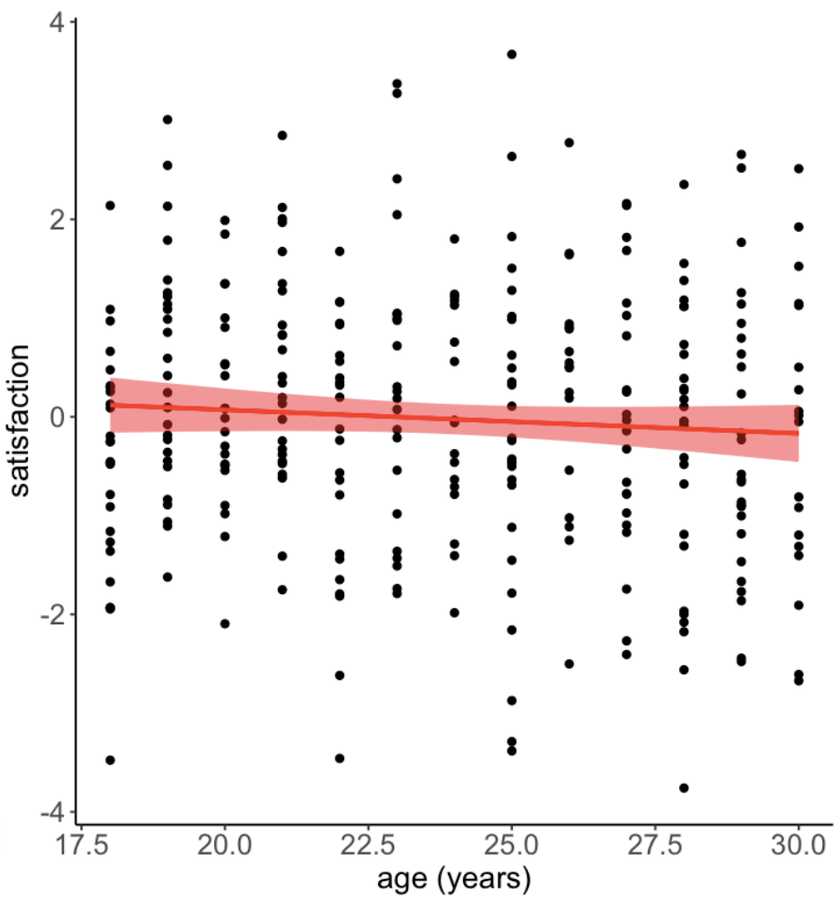
satisfaction	age	exam1
.05	20	75
1.23	30	80
2.9	18	96
.30	23	92
.99	21	73

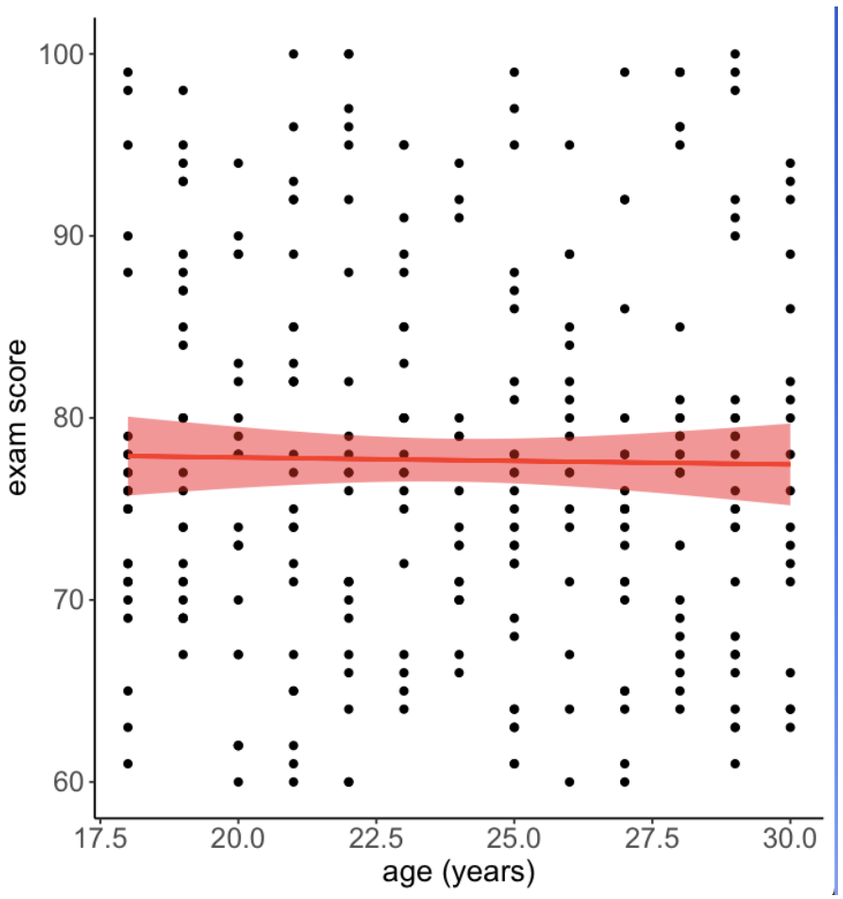
correlation

assumptions

- confirm that the relationship between the DV and IV are linear







hypothesis for correlation

- H0: x variable and y variable are not correlated (p = 0).
- HA: x variable and y variable are correlated (p ≠ 0).

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2
1. satisfaction	-0.02	1.34		
2. exam1	77.68	10.38	.56** [.47, .63]	
3. age	23.85	3.81	-.07 [-.18, .05]	-.01 [-.13, .10]

interpretation for correlation

- This data provides evidence that **exam 1 score is significantly correlated to class satisfaction (t(298) = 11.60, p < .001)**. There is a **strong positive** relationship between score and satisfaction, and we are **95% confident** that the correlation coefficient is between **.47 and .63**.

- We found no evidence that the relationship between age and exam 1 score is correlated (t(298) = -.24, p = .80)

Table 1

Means, standard deviations, and correlations with confidence intervals

Variable	*M*	*SD*	1	2
1. satisfaction	-0.02	1.34		
2. exam1	77.68	10.38	.56** [.47, .63]	
3. age	23.85	3.81	-.07 [-.18, .05]	-.01 [-.13, .10]

regression

assumptions

- confirm that the relationship between the DV and IV are linear
- normality
  - qqplots
  - residual plots

hypothesis for regression

- H0: x variable is **not** a linear predictor of y variable ($\beta = 0$).
- HA: x variable is a linear predictor of y variable ($\beta \neq 0$).

interpretation for regression

- The data shows evidence that class satisfaction is **significantly linearly** related to exam 1 score **(t(298) = 11.6, p < .001)**. For every additional point increase in satisfaction, **exam 1 score increased by 4.32, on average.**
- There was not a significantly linear relationship between age and exam 1 score (t(298) = -.24, p = .81).

general linear model

assumptions

- ✓ confirm that the relationship between the DV and IV are linear
- ✓ normality
  - qqplots
  - residual plots for equal variance

hypothesis for glm

- exam 1 score ~ satisfaction + time
- H0: when controlling for class time, satisfaction is **not** a linear predictor of exam 1 score.
- HA: when controlling for class time, satisfaction is a linear predictor of exam 1 score.
- H0: when controlling for satisfaction, class time is **not** a linear predictor of exam 1 score.
- HA: when controlling for satisfaction, class time is a linear predictor of exam 1 score.

hypothesis for glm

- exam 1 score ~ satisfaction + time
- H0: when controlling for class time, satisfaction is **not** a linear predictor of exam 1 score.
- HA: when controlling for class time, satisfaction is a linear predictor of exam 1 score.
- H0: when controlling for satisfaction, class time is **not** a linear predictor of exam 1 score.
- HA: when controlling for satisfaction, class time is a linear predictor of exam 1 score.

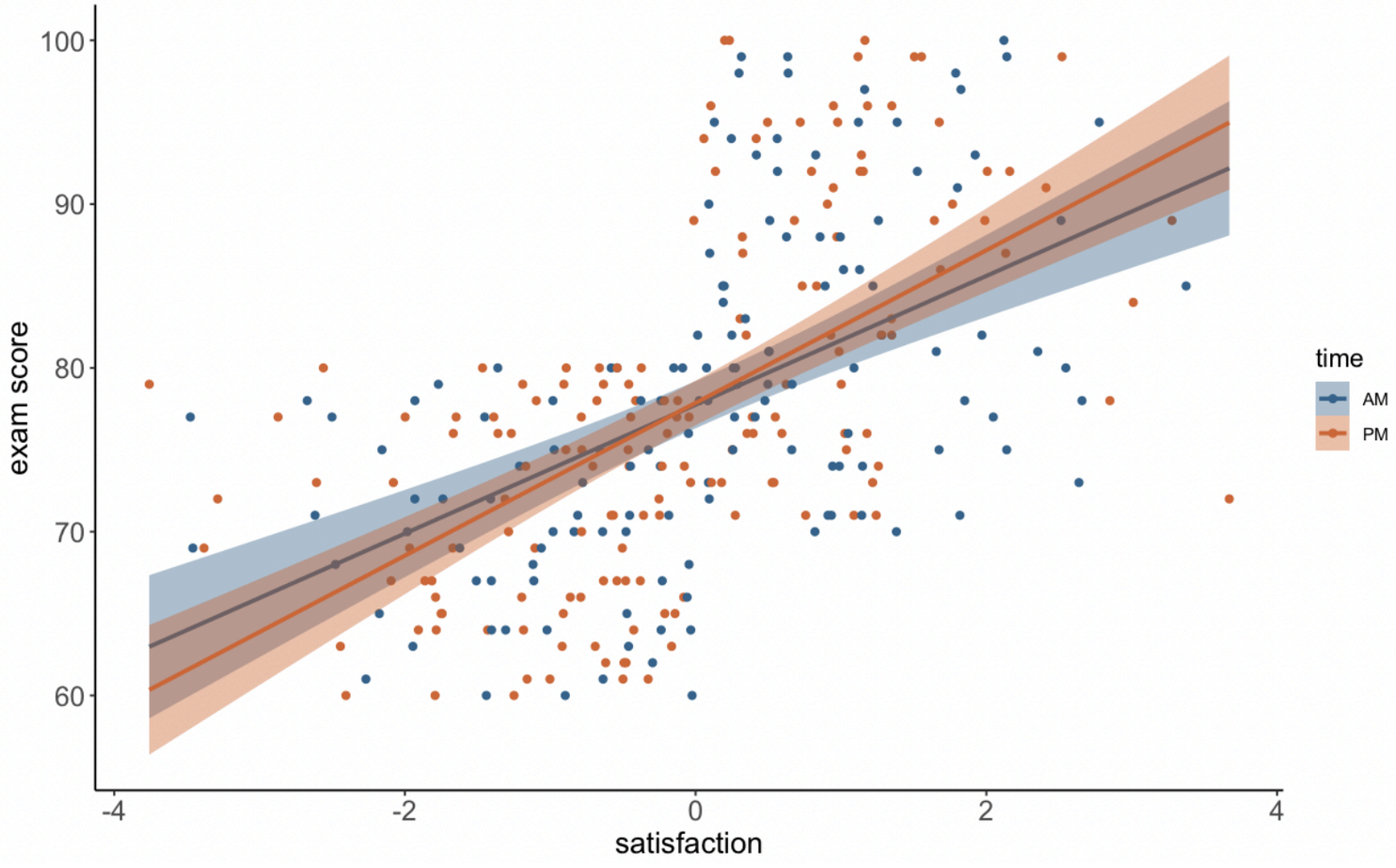
hypothesis for glm

- exam 1 score ~ satisfaction + time
- H0: when controlling for class time, satisfaction is **not** a linear predictor of exam 1 score.
- HA: when controlling for class time, satisfaction is a linear predictor of exam 1 score.
- H0: when controlling for satisfaction, class time is **not** a linear predictor of exam 1 score.
- HA: when controlling for satisfaction, class time is a linear predictor of exam 1 score.

interpretation for glm

- While controlling for class time, satisfaction is a **significant linear predictor** of exam 1 score **(t(2, 297) = 11.55, p < .001)**. For every one point increase in satisfaction, there was a **4.32 points increase** in exam score, **on average while holding class time constant.**

- While controlling for satisfaction, class time was not a significant linear predictor of exam 1 score (t(2, 297) = .10, p = .92).



interaction for glm

- similar to multi-factor ANOVA, we can test if the effect of one variable differs based on the other
- H0: the effect of satisfaction on exam 1 score is the same for AM and PM classes.
- HA: the effect of satisfaction on exam 1 score is **not** the same for AM and PM classes. (there is an interaction between satisfaction x time on exam 1)

interaction for glm

- similar to multi-factor ANOVA, we can test if the effect of one variable differs based on the other
- H0: the effect of satisfaction on exam 1 score is the same for AM and PM classes.
- HA: the effect of satisfaction on exam 1 score is **not** the same for AM and PM classes. (there is an interaction between satisfaction x time on exam 1)

- if there was a sign. interaction
    - There was a significant interaction between class time and satisfaction on exam 1 score. We found that students in the afternoon class had a positive linear relationship between satisfaction and exam 1 score (insert stats), whereas the students in the morning class had a non-significant linear relationship between satisfaction and exam score (insert stats).