Name: KEY

MATH 321: Review Part 3

1. The University is investigating the safety of MATH courses for their students. To do this, they plan a study to compare the blood pressure of upper level MATH courses students before the final exam and after completing the final exam. Data are shown below (in mm Hg).

   Construct and interpret a 95% confidence interval for the difference in blood pressure before and after the final exam for math students. State a conclusion if final exams increased blood pressure on average.

BP Before	BP After	Before - After
120	125	-5
110	112	-2
208	207	1
121	128	-7
115	119	-4
119	115	4
123	128	-5
117	125	-8
124	124	0

Small n & dependent samples $\Rightarrow$ paired t interval on differences

technically need to look at histogram of differences to confirm $\approx$ approx normal

1-var stats (differences) $\rightarrow$ $\bar{x} = -2.888$, $s_x = 3.951$

OR L_1 = Before, L_2 = After, L_3: $L_1 - L_2$

$$95\%\ CI \approx \bar{d} \pm t_{8,0.025}\, s_D/\sqrt{n} = -2.888 \pm 2.306\left(\frac{3.951}{\sqrt{9}}\right) = \boxed{[-5.925,\ 0.149]}$$

contains 0

$\Rightarrow$ $\boxed{\text{No conclusion if exams raised BP}}$

Interpretation $\rightarrow$ We are 95% confident the true mean difference in blood pressure after & before the exam is between -5.925 & 0.149

2. The campus bookstore is determining if they need to increase their marketing budget. They would like at least 65% of students to buy their textbooks directly from them rather than off-campus stores. In order to check this, they took a random sample of 137 students in which 81 students said they buy their books at the campus bookstore.

   (a) Is there enough evidence to conclude the true proportion of students who buy their books at the campus bookstore is greater than 65%? Use $\alpha = 0.05$.

$np_0 = 137(0.65) = 89.05 \geq 5$ ✓

$n(1-p_0) = 137(0.35) = 47.95 \geq 5$ ✓

$\Rightarrow$ Conditions met ✓

$\rightarrow$ Let p = true proportion of students who buy textbooks at the campus bookstore.

$\rightarrow$ $H_0: p = 0.65$

$H_a: p > 0.65$

$\rightarrow$ TS: $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \dfrac{\frac{81}{137} - 0.65}{\sqrt{\frac{0.65 \cdot 0.35}{137}}} = -1.442$ $\longrightarrow$ p-value: $P(z > -1.442) = 0.9253 > \alpha \Rightarrow$ FTR X

$\rightarrow$ RR: $\{z > z_{0.05}\} = \{z > 1.645\} \Rightarrow$ fail to reject X

invNorm(0.05, 0, 1)

$\rightarrow$ Conclusion: At the 5% significance level, there is not sufficient evidence to conclude the true proportion of students who purchase textbooks at the campus bookstore is greater than 0.65

(b) Construct the corresponding 90% (two-sided) confidence interval. State a conclusion if this proportion is greater at least the desired 65%.

$$90\%\ CI \approx \hat{p} \pm z_{0.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \overset{81/137}{0.591} \pm 1.645\sqrt{\frac{0.591(0.409)}{137}} \boxed{= [0.522, 0.660]}$$

Contains 0.65 $\Rightarrow$ $\boxed{\text{not} > 0.65}$

3. Scientists discovered a new mountain range under the sea. Lets assume the sea mountain heights are normally distributed with unknown standard deviation.

From a random sample of 13 peaks, there was an average height of 11,308 ft and standard deviation of 5,287 ft.

(a) Is there enough evidence to conclude the average heights of these new sea mountains is different than the Rocky Mountains, which average 14,400 ft? Use $\alpha = 0.12$.

$\rightarrow$ Let μ = true mean height of under-sea mountain range (ft)

$X \sim N(\mu, \sigma^2 = ??)$

small n

$\Rightarrow t$

$H_0: \mu = 14{,}400$

$H_A: \mu \neq 14{,}400$

TS: $T = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{11308 - 14400}{5287/\sqrt{13}} = -2.109 \longrightarrow$ p-value $= 2 \cdot P(|t_{12}| > |-2.109|)$

$= 2 \cdot 0.028$

$= 0.056 < \alpha \Rightarrow$ Reject ✓

RR: $\{|t| > t_{0.06,12}\} = \{|t| > 1.674\} \Rightarrow$ Reject H_0 ✓

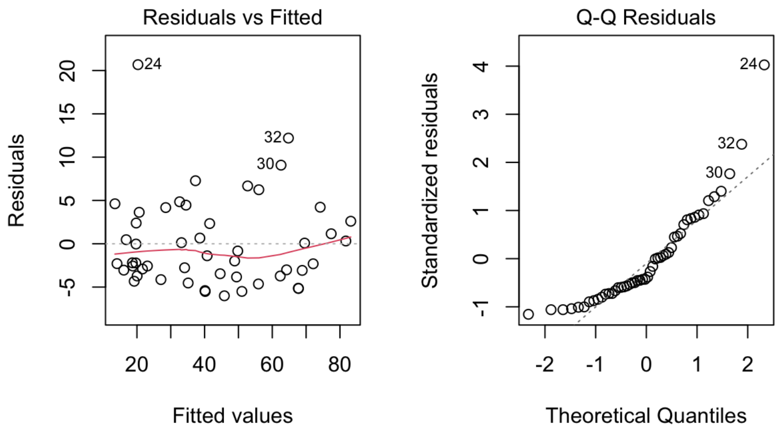
Conclusion: At $\alpha = 0.12$, there is sufficient evidence to conclude that the true mean height of these new mountains $\neq$ 14,400 ft, it is actually smaller.

(b) Construct the corresponding (two-sided) confidence interval for this test and confirm your conclusion from part (a).

$$88\%\ CI \approx \bar{x} \pm t_{12,0.06}\frac{s}{\sqrt{n}} = 11308 \pm 1.674 \frac{5287}{\sqrt{13}} = [8853.33, 13762.67] < 14{,}400 \Rightarrow \text{Reject ✓}$$

4. Given the following residual plots for the regression of $Y \sim X$, assess the assumptions of the normal error regression model.

↳ LINE



→ Linearity → No systematic deviations in the residuals vs fitted plot ⟹ no issues

→ Independence → Assuming data was collected in a randomized experiment ⟹ no issues

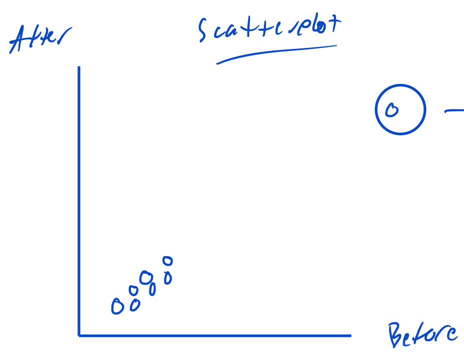
→ Normality → Residuals do not follow the trend line ⟹ minor concerns about this

→ Equal variance → No tipped-over-tornado in the residuals vs fitted plot ⟹ no issues

⟹ only concern is normality

5. (Extra R practice, NOT ON FINAL)

   (a) Using the data from question 1, perform EDA for the regression on After as the response variable and Before as the explanatory variable. Is there anything we should be concerned about beforehand?



→ outlier may cause issues, consider removing

   (b) Fit the regression from part (a) and get the estimated coefficients.

lm(After ~ Before) ⟹ $\widehat{After} = 9.0706 + 0.9519$ Before

(c) Perform a t-test on the slope coefficient. Use $\alpha = 0.01$

→ Let β_1 = True population slope for After ~ Before

→ H_0: $\beta_1 = 0$

H_A: $\beta_1 \neq 0$

→ TS: $t = \frac{\hat{\beta}_1 - 0}{S_{\hat{\beta}_1}} = \frac{0.9519 - 0}{0.046} = 20.617$ → p-value $= 2 \cdot P(|t_?| > 20.617) \approx 0$ ⇒ reject ✓

→ Conclusion: At the 1% significance level, there is sufficient evidence to conclude the true population slope for Blood pressure before ≠ 0

(d) Predict the after blood pressure for a student with a before blood pressure of 123.

$\widehat{After} = 9.0706 + 0.9519\ (123)$

$\boxed{\downarrow = 126.1543}$ ⟶ predict(mod, newdata = data.frame(before = 123)