GOODNESS OF FIT

How to tell it's Goodness of Fit
• There is only 1 variable
• There are old/known values & new observed values

Hypotheses			
H_0	The proportions are correct.	H_1	At least 1 proportion is incorrect.

Math Time!	
Expected Counts	$E_i = n \cdot p_i$
Test Statistic	$\chi^2 = TS = \sum \left(\frac{(O_i - E_i)^2}{E_i} \right)$
Degrees of Freedom	$df = \#\text{categories} - 1$
P-value	$p = \chi^2 cdf(TS, E99, df)$

Decision and Interpretation	
Rejecting Null ($\alpha > p$)	There is sufficient evidence to say that at least 1 proportion is incorrect.
Failing to Reject Null ($\alpha < p$)	There is insufficient evidence to say that at least 1 proportion is incorrect.

Calculator Trick (X^2GOF-Test)

How do you get there?	STAT → TESTS → D: χ^2GOF-Test
What do you need for it?	• Observed Counts in list • Expected Counts in list • Degrees of Freedom
What do you get from it?	• TS → χ^2 • p-value • Df

TEST OF INDEPENDENCE & TEST OF HOMOGENEITY

How to tell it's...	
Test of Independence	**Test of Homogeneity**
• There are 2 variables • Question has words like 'relationship' or 'association'	• There are 2 variables • Question has phrases like 'the same' or 'are equal'

Hypotheses			
Test of Independence		**Test of Homogeneity**	
H_0	The variables are independent	H_0	The distributions are equal/the same
H_1	The variables are dependent	H_1	The distributions are unequal/not the same

Math Time!	
Find Row/Column Totals & Overall Total	*Note: Overall total is found by adding row totals **OR** column totals
Expected Counts	$E_{i,j} = \frac{(i^{th} \text{ row total})(j^{th} \text{ column total})}{\text{Overall total}}$
Test Statistic	$\chi^2 = TS = \sum\left(\frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}\right)$

Degrees of Freedom	df = (#rows − 1)(#columns − 1)
P-value	$p = \chi^2cdf(TS, E99, df)$

Decision and Interpretation		
Rejecting Null ($\alpha > p$)	TOI	There is sufficient evidence to say that the variables are dependent of each other.
	TOH	There is sufficient evidence to say that the distributions are unequal.
Failing to Reject Null ($\alpha < p$)	TOI	There is insufficient evidence to say that the variables are dependent of each other.
	TOH	There is insufficient evidence to say that the distributions are unequal.

Calculator Trick	
Part 1: The Matrix	
How do you get there?	2nd → x^{-1} → Edit → Pick one
What do you need for it?	Observed Counts
Part 2: X^2-Test	
How do you get there?	STAT → TESTS → C: χ^2–Test
What do you need for it?	• Observed Count Matrix • Empty Matrix for Expected Counts
What do you get from it?	• Expected Counts • $TS = \chi^2$ • p-value • df

ANOVA

How to tell it's ANOVA
• There is only 1 variable • Question often has phrases like 'most', 'least', or 'different' when discussing the levels of the variable

Hypotheses			
H_0	$\mu_1 = \mu_2 = \cdots = \mu_T$	H_1	At least 1 mean differs

Math Time!		
	Between	Within
Sum of Squares	$SSB = \sum(n_i(\bar{x}_i - \bar{x})^2)$	$SSW = \sum((n_i - 1)s_i^2)$
Degrees of Freedom	$df_B = T - 1$	$df_W = n - T$
Mean of Squares	$MSB = \frac{SSB}{df_B}$	$MSW = \frac{SSW}{df_W}$
Test Statistic	$F = TS = \frac{MSB}{MSW}$	
P-value	$p = Fcdf(TS, E99, Df_B, Df_W)$	

Decision and Interpretation	
Rejecting Null ($\alpha > p$)	There is sufficient evidence to say that at least 1 mean differs.

Failing to Reject Null ($\alpha < p$)	There is insufficient evidence to say that at least 1 mean differs.

Calculator Trick (ANOVA)	
How do you get there?	STAT → TESTS → H:ANOVA(
What do you need for it?	All the levels' data in lists
What do you get from it?	• F=TS • p-value • Factor Stats → Between Values • Error Stats → Within Values

POST-HOC

How to tell it's Post-Hoc
• Question asks something along the line of which mean differs from the rest *Note: Only use when you REJECT NULL on ANOVA!

Suppose there are 3 μ

Hypotheses			
H_0	$\mu_1=\mu_2$; $\mu_1=\mu_3$; $\mu_2=\mu_3$	H_1	$\mu_1\neq\mu_2$; $\mu_1\neq\mu_3$; $\mu_2\neq\mu_3$

Decision and Interpretation for Each Comparison	
Rejecting Null ($\alpha > p$)	$\mu_{\#} \neq \mu_{\#}$
Failing to Reject Null ($\alpha < p$)	$\mu_{\#} = \mu_{\#}$

Interpretation Overall
There is sufficient evidence to say that the mean of ____________ are different from ___________.   * Honestly, just write/chose what you see in your analysis of the R code you are provided.