



- Quizzes Review Test Submission: MBC638 Quiz #12 - Hypothesis testing (due Sunday, Dec. 2, 10:00pm)

Review Test Submission: MBC638 Quiz #12 - Hypothesis testing (due Sunday, Dec. 2, 10:00pm)

User	David Forteguerre
Course	MBC.638.M001.FALL18.Data Anls & Decisn Making
Test	MBC638 Quiz #12 - Hypothesis testing (due Sunday, Dec. 2, 10:00pm)
Started	12/1/18 9:28 PM
Submitted	12/1/18 10:10 PM
Status	Completed
Attempt Score	100 out of 100 points
Time Elapsed	41 minutes out of 1 hour

Results Displayed All Answers, Submitted Answers, Feedback, Incorrectly Answered Questions

Question 1

5 out of 5 points



Table A. Occupational employment and wages by major occupational group, United States and the San Jose-Sunnyvale-Santa Clara Metropolitan Statistical Area, and measures of statistical significance, May 2016

Major occupational group	Percent of total employment		Mean hourly wage		Percent difference <small>(1)</small>
	United States	San Jose	United States	San Jose	
Total, all occupations	100.0	100.0	\$23.86	\$37.98*	59
Management	5.1	8.1**	56.74	80.09*	41
Business and financial operations	5.2	7.3**	36.09	48.78*	35
Computer and mathematical	3.0	12.5**	42.25	59.42*	41
Architecture and engineering	1.8	5.4**	40.53	55.29*	36
Life, physical, and social science	0.8	1.1**	35.08	43.78*	25
Community and social service	1.4	1.0**	22.69	27.32*	20
Legal	0.8	0.7	50.95	75.53*	48
Education, training, and library	6.2	5.6**	28.21	31.98*	22
Arts, design, entertainment, sports, and media	1.4	1.4	28.07	34.29*	22
Healthcare practitioners and technical	5.9	4.1**	38.08	54.57*	43
Healthcare support	2.9	1.7**	14.85	20.33*	39
Protective service	2.4	1.7**	22.03	29.39*	33
Food preparation and serving related	9.2	7.5**	11.47	13.98*	22
Building and grounds cleaning and maintenance	3.2	2.7**	13.47	15.98*	18
Personal care and service	3.2	2.6**	12.74	14.78*	16
Sales and related	10.4	9.2**	19.50	27.13*	39
Office and administrative support	15.7	12.5**	17.91	23.58*	32
Farming, fishing, and forestry	0.3	0.3	13.37	11.90*	-11
Construction and extraction	4.0	3.5**	23.51	30.80*	31
Installation, maintenance, and repair	3.9	2.6**	22.45	27.54*	23
Production	6.5	5.0**	17.88	20.37*	14
Transportation and material moving	6.9	3.7**	17.34	18.38*	6

According to the Bureau of Labor Statistics, the hourly wages in the "business and financial operations" industry averaged \$36.09 in 2016. A sample of 1,000 workers in the San Jose-Sunnyvale-Santa Clara Metropolitan Statistical Area showed an average hourly wage of \$49.52 with a standard deviation of \$12.01.

Does the data provide sufficient evidence to conclude that the average hourly wage in the business and financial operations industry in the San Jose-Sunnyvale-Santa Clara region is above the national average? From the list below, pick an appropriate set of null and alternative hypotheses.

Selected Answer: $H_0: \mu \leq 36.09$

$H_A: \mu > 36.09$

Answers: $H_0: \mu \geq 36.09$

$H_A: \mu < 36.09$

$H_0: \mu \geq 49.52$

$H_A: \mu < 49.52$

$H_0: \mu = 36.09$

$H_A: \mu \neq 36.09$

$$H_0: \mu = 49.52$$

$$H_A: \mu \neq 49.52$$

$$H_0: \mu \leq 36.09$$

$$H_A: \mu > 36.09$$

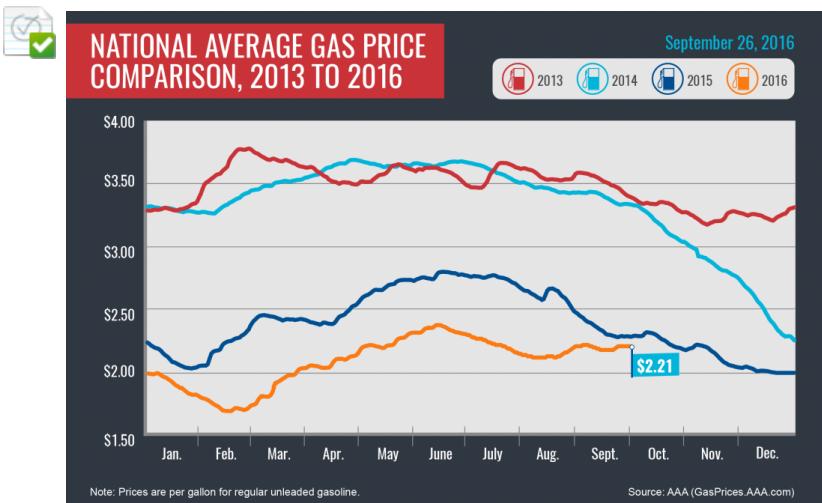
$$H_0: \mu \leq 49.52$$

$$H_A: \mu > 49.52$$

Response Feedback:

Question 2

10 out of 10 points



According to AAA, the national average gas price on September 26, 2016 was \$2.21 per gallon for regular unleaded gasoline. Assume that gas prices follow the Normal distribution. A small survey was conducted and the gas prices for regular unleaded gasoline from 10 gas stations' from different cities nationwide were recorded exactly one year later, on September 26, 2017. Their average was found to be \$2.41 per gallon with a standard deviation of \$0.58. Does the sample provide sufficient evidence that the average gas prices have risen over the one year period?

STEP 1. State the appropriate null and alternative hypotheses. # [a]

#1: $H_0: \mu \geq 2.21 \quad H_A: \mu < 2.21$ (left-tailed test)

#2: $H_0: \mu \leq 2.21 \quad H_A: \mu > 2.21$ (right-tailed test)

#3: $H_0: \mu = 2.21 \quad H_A: \mu \neq 2.21$ (two-tailed test)

STEP 2. Compute the test statistic for this test. Test statistic = [b] + [c][d][e] (round to 3 decimal places)

STEP 3: Compute the p-value. p-value = 0 + [f][g][h][i]

STEP 4: Using $\alpha = 5\%$, state your conclusions:

Is the null hypothesis rejected? (0=NO, 1=YES) [j]

Pick the correct statement. # [k]

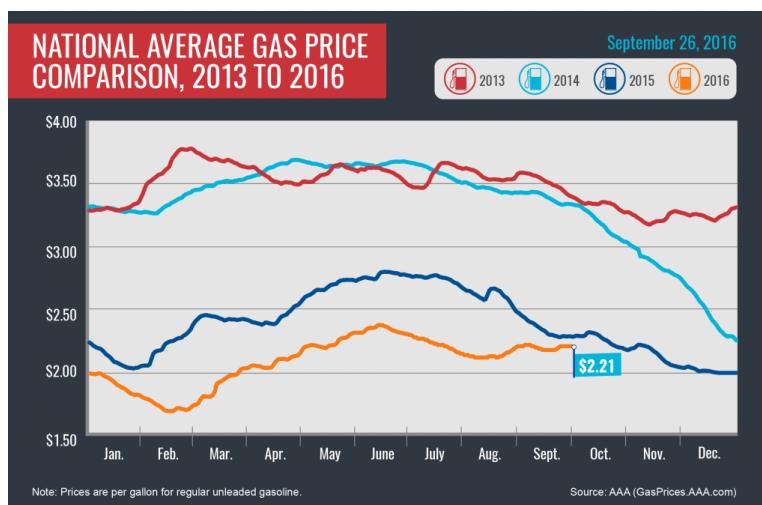
#1: Yes, we have sufficient evidence that in this sample the average price is higher than last year's \$2.21.

#2: No, we don't have sufficient evidence that in this sample the average price is higher than last year's \$2.21.

#3: Yes, we have sufficient evidence that the average price in 2017 nationally is higher than last year's \$2.21.

#4: No, we don't have sufficient evidence that the average price in 2017 nationally is higher than last year's \$2.21.

Selected Answer:



According to AAA, the national average gas price on September 26, 2016 was \$2.21 per gallon for regular unleaded gasoline. Assume that gas prices follow the Normal distribution. A small survey was conducted and the gas prices for regular unleaded gasoline from 10 gas stations' from different cities nationwide were recorded exactly one year later, on September 26, 2017. Their average was found to be \$2.41 per gallon with a standard deviation of \$0.58. Does the sample provide sufficient evidence that the average gas prices have risen over the one year period?

STEP 1. State the appropriate null and alternative hypotheses. **# 2**

#1: $H_0: \mu \geq 2.21$ $H_A: \mu < 2.21$ (left-tailed test)

#2: $H_0: \mu \leq 2.21$ $H_A: \mu > 2.21$ (right-tailed test)

#3: $H_0: \mu = 2.21$ $H_A: \mu \neq 2.21$ (two-tailed test)

STEP 2. Compute the test statistic for this test. Test statistic = 1 . 090 (round to 3 decimal places)

STEP 3: Compute the p-value. p-value = 0 . 1519

STEP 4: Using $\alpha = 5\%$, state your conclusions:

Is the null hypothesis rejected? (0=NO, 1=YES) 0

Pick the correct statement. **# 4**

#1: Yes, we have sufficient evidence that in this sample the average price is higher than last year's \$2.21.

#2: No, we don't have sufficient evidence that in this sample the average price is higher than last year's \$2.21.

#3: Yes, we have sufficient evidence that the average price in 2017 nationally is higher than last year's \$2.21.

#4: No, we don't have sufficient evidence that the average price in 2017 nationally is higher than last year's \$2.21.

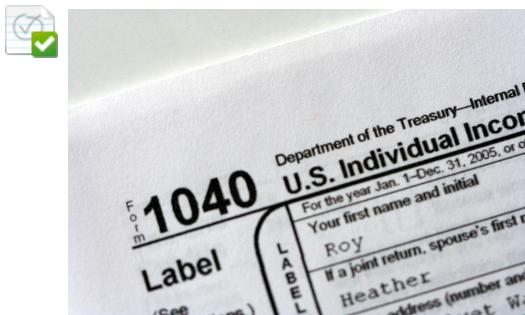
Answers:

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback: 😊

Question 3

10 out of 10 points

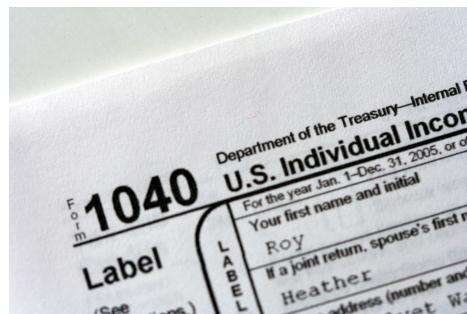


The IRS would like to estimate the total net amount of refund. Each taxpayer will either receive a refund, in which case the net refund is positive, or will have to pay an amount due, in which case the net refund is negative.

The data [IRS Refunds.xlsx](#) contains net refund amounts for 500 random taxpayers. Based on this sample data, an IRS analyst believes that the national average net refund amount exceeds \$250. Does the data provide evidence to support this?

- The **test statistic** associated with this test equals [a] + [c][d][e][f] (Round to 4 decimal places)
- **p-value** equals 0 + [g][h][i][j] (Round to 4 decimal places) (Let's agree that whenever σ is known we use Z, and whenever σ is unknown we use t)
- **At $\alpha=5\%$, what is the conclusion? # [k]**
 - **#1:** Yes, the data provides sufficient evidence that the national average net refund amount exceeds \$250.
 - **#2:** No, the data doesn't provide sufficient evidence that the national average net refund amount exceeds \$250.

Selected
Answer:



The IRS would like to estimate the total net amount of refund. Each taxpayer will either receive a refund, in which case the net refund is positive, or will have to pay an amount due, in which case the net refund is negative.

The data [IRS Refunds.xlsx](#) contains net refund amounts for 500 random taxpayers. Based on this sample data, an IRS analyst believes that the national average net refund amount exceeds \$250. Does the data provide evidence to support this?

- The **test statistic** associated with this test equals **1 . 7302** (Round to 4 decimal places)
- **p-value** equals **0 . 0421** (Round to 4 decimal places) (Let's agree that whenever σ is known we use Z, and whenever σ is unknown we use t)
- At **$\alpha=5\%$, what is the conclusion? # 1**
 - **#1:** Yes, the data provides sufficient evidence that the national average net refund amount exceeds \$250.
 - **#2:** No, the data doesn't provide sufficient evidence that the national average net refund amount exceeds \$250.

Answers:

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback:

Question 4

15 out of 15 points



McDonald's is a restaurant system that is famous for providing food that tastes the same in Alaska as it does in New York.

Every now and then, McDonald's re-assesses their food options and tries to come up with healthier alternatives. How many calories does the average McDonald's value meal contain? How much do beverages, like soda or coffee, contribute to the overall caloric intake? Does ordered grilled chicken instead of crispy increase a sandwich's nutritional value?

In this problem, you are asked to examine the menu items and nutrition facts that were obtained from the McDonald's website.

Use Excel file [McDonald menu.xlsx](#). This dataset contains information on the menu items and nutrition facts obtained from the McDonald's website. The menu constantly changes, so you can think of this dataset as a sample.

We are interested in the **average** level of **Calories** (column D) for McDonald's food items.

a) A competitor Burger King has released a new TV commercial in which it claims that the amount of calories in McDonald's food exceeds 400 on average. You have done a summer internship at McDonald's and believe that this figure is overstated; therefore, you want to refute the claim (i.e., prove him wrong) made in the commercial.

- Form an appropriate hypothesis. Select one from the following three options. # [a]

#1: $H_0: \mu \leq 400$
 $H_A: \mu > 400$

#2: $H_0: \mu \geq 400$
 $H_A: \mu < 400$

#3: $H_0: \mu = 400$
 $H_A: \mu \neq 400$

- What is the value of the **test statistic?** (Round to 4 decimal places.) – [b] . [c][d][e][f]
- What is the **p-value?** (Round to 4 decimal places.) 0 . [cc][dd][ee][ff]

- Using $\alpha=0.01$, can you credibly refute the claim about McDonald's food calorie content made in Burger King's TV commercial? (0=No, 1=Yes) [g]

b) You would like to analyze and compare the calorie content for "Breakfast" and all other food categories (column A). Your hypothesis is that the average calorie content differs for breakfast and all other food categories.

- Form an appropriate hypothesis. Select one from the following three options. # [aa]

#1: $H_0: \mu_1 \leq \mu_2$
 $H_A: \mu_1 > \mu_2$

#2: $H_0: \mu_1 \geq \mu_2$
 $H_A: \mu_1 < \mu_2$

#3: $H_0: \mu_1 = \mu_2$
 $H_A: \mu_1 \neq \mu_2$

- What is the absolute value of the **test statistic**? (Perform the analysis in Excel, either StatTools or Data Analysis Toolpak. Round to 4 decimal places.) [h] . [i][j][k][l]
- Using $\alpha=0.01$, can you conclude that the average calorie content is different for breakfast and all other food categories? (0=No, 1=Yes) [m]

Selected

Answer:



McDonald's is a restaurant system that is famous for providing food that tastes the same in Alaska as it does in New York.

Every now and then, McDonald's re-assesses their food options and tries to come up with healthier alternatives. How many calories does the average McDonald's value meal contain? How much do beverages, like soda or coffee, contribute to the overall caloric intake? Does ordered grilled chicken instead of crispy increase a sandwich's nutritional value?

In this problem, you are asked to examine the menu items and nutrition facts that were obtained from the McDonald's website.

Use Excel file [McDonald menu.xlsx](#). This dataset contains information on the menu items and nutrition facts obtained from the McDonald's website. The menu constantly changes, so you can think of this dataset as a sample.

We are interested in the **average** level of **Calories** (column D) for McDonald's food items.

a) A competitor Burger King has released a new TV commercial in which it claims that the amount of calories in McDonald's food exceeds 400 on average. You have done a summer internship at McDonald's and believe that this figure is overstated; therefore, you want to refute the claim (i.e., prove him wrong) made in the commercial.

- Form an appropriate hypothesis. Select one from the following three options. # 2

#1: $H_0: \mu \leq 400$
 $H_A: \mu > 400$

#2: $H_0: \mu \geq 400$
 $H_A: \mu < 400$

#3: $H_0: \mu = 400$
 $H_A: \mu \neq 400$

- What is the value of the **test statistic**? (Round to 4 decimal places.) – 2 . 1295
- What is the **p-value**? (Round to 4 decimal places.) 0 . 0171
- Using $\alpha=0.01$, can you credibly refute the claim about McDonald's food calorie content made in Burger King's TV commercial? (0=No, 1=Yes) 0

b) You would like to analyze and compare the calorie content for "Breakfast" and all other food categories (column A). Your hypothesis is that the average calorie content differs for breakfast and all other food categories.

- Form an appropriate hypothesis. Select one from the following three options. # 3

#1: $H_0: \mu_1 \leq \mu_2$
 $H_A: \mu_1 > \mu_2$

#2: $H_0: \mu_1 \geq \mu_2$
 $H_A: \mu_1 < \mu_2$

#3: $H_0: \mu_1 = \mu_2$
 $H_A: \mu_1 \neq \mu_2$

- What is the absolute value of the **test statistic**? (Perform the analysis in Excel, either StatTools or Data Analysis Toolpak. Round to 4 decimal places.) 5 . 0187

- Using $\alpha=0.01$, can you conclude that the average calorie content is different for breakfast and all other food categories? (0=No, 1=Yes) **1**

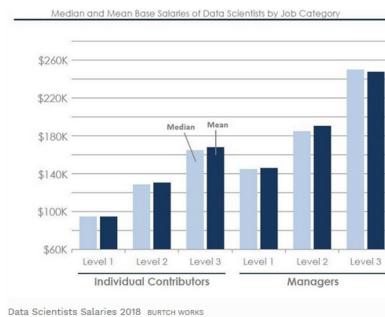
Answers:

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback:

Question 5

15 out of 15 points



According to [Forbes \(May 4, 2018\)](#), data scientists are in high demand and their salaries have been growing. Is getting a degree in Applied Data Science worth the money? Let's look at some data.

A sample of 10 recent graduates with a Master's degree in Applied Data Science were surveyed and their salaries right before enrolling in the program and their starting salary right after obtaining the degree were recorded. Assume that starting salaries follow a bell-shaped distribution. The following table summarizes the data. Let's agree that whenever σ is known we use Z and whenever σ is unknown we use t.

Name	Salary (\$ '000)	
	Before	After
Xinxia	55	95
Deep	49	100
Anukrati	36	180
Qiaomu	77	88
David	85	125
Xuehan	49	91
Yunxia	46	185
Harsh	79	103
Nicholas	99	220
Ananth	110	110

(Hint: Copy and paste this table into Excel.)

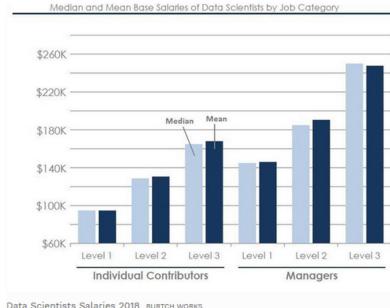
Formulate an appropriate hypothesis if your objective is to determine whether a degree in Applied Data Science helps boost your salary.

- What is the absolute value of the **test statistic**? (Round to 6 decimal places.) [a] . [b][c][d][e][f][g]
- What is the **p-value**? (Round to 6 decimal places.) 0 . [bb][cc][dd][ee][ff][gg]
- State your conclusions using $\alpha=5\%$: # [h]

#1: Yes, we have sufficient evidence that having a degree in Applied Data Science helps boost average salary.

#2: No, we don't have sufficient evidence that having a degree in Applied Data Science helps boost average salary.

Selected Answer:



According to [Forbes \(May 4, 2018\)](#), data scientists are in high demand and their salaries have been growing. Is getting a degree in Applied Data Science worth the money? Let's look at some data.

A sample of 10 recent graduates with a Master's degree in Applied Data Science were surveyed and their salaries right before enrolling in the program and their starting salary right after obtaining the degree were recorded. Assume that starting salaries follow a bell-shaped distribution. The following table summarizes the data. Let's agree that whenever σ is known we use Z and whenever σ is unknown we use t.

Name	Salary (\$ '000)	
	Before	After
Xinxia	55	95
Deep	49	100
Anukrati	36	180
Qiaomu	77	88
David	85	125
Xuehan	49	91
Yunxia	46	185
Harsh	79	103
Nicholas	99	220
Ananth	110	110

(Hint: Copy and paste this table into Excel.)

Formulate an appropriate hypothesis if your objective is to determine whether a degree in Applied Data Science helps boost your salary.

- What is the absolute value of the **test statistic**? (Round to 6 decimal places.) 3 . 635423
- What is the **p-value**? (Round to 6 decimal places.) 0 . 002720
- State your conclusions using $\alpha=5\%$: # 1

#1: Yes, we have sufficient evidence that having a degree in Applied Data Science helps boost average salary.

#2: No, we don't have sufficient evidence that having a degree in Applied Data Science helps boost average salary.

Answers:

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback: 😊

Question 6

15 out of 15 points



According to the [Fortune Magazine](#), MBA graduates were getting an average of \$45,000 raise in 2015.

A sample of 10 recent MBA graduates were surveyed and their salaries right before enrolling in an MBA program and their starting salary right after obtaining the degree were recorded. Assume that starting salaries follow a bell-shaped distribution. The following table summarizes the data. Let's agree that whenever σ is known we use Z and whenever σ is unknown we use t.

Name	Salary (\$ '000)	
	Before MBA	After MBA
Minyang	32	300
Palaniappan	80	105
Sijie	35	100
Apurva	74	99
Spiti	49	90
Yun	59	120
Eshank	36	140
Xueqing	90	115
Yesaswi	100	350
Rahul	25	250

(Hint: Copy and paste this table into Excel.)

Formulate an appropriate hypothesis if your objective is to determine whether an MBA degree helps boost your salary **by more than \$45,000 on average**.

- What is the absolute value of the **test statistic**? (Round to 6 decimal places.) [a] . [b][c][d][e][f][g]
- What is the **p-value**? (Round to 6 decimal places.) 0 . [bb][cc][dd][ee][ff][gg]
- State your conclusions using $\alpha=5\%$: # [h]

#1: Yes, we have sufficient evidence that having an MBA degree helps boost average salary by more than \$45,000 on average.

#2: No, we don't have sufficient evidence that having an MBA degree helps boost average salary by more than \$45,000 on average.

Selected Answer: According to the [Fortune Magazine](#), MBA graduates were getting an average of

\$45,000 raise in 2015.

A sample of 10 recent MBA graduates were surveyed and their salaries right before enrolling in an MBA program and their starting salary right after obtaining the degree were recorded. Assume that starting salaries follow a bell-shaped distribution. The following table summarizes the data. Let's agree that whenever σ is known we use Z and whenever σ is unknown we use t.

Name	Salary (\$ '000)	
	Before MBA	After MBA
Minyang	32	300
Palaniappan	80	105
Sijie	35	100
Apurva	74	99
Spiti	49	90
Yun	59	120
Eshank	36	140
Xueqing	90	115
Yesaswi	100	350
Rahul	25	250

(Hint: Copy and paste this table into Excel.)

Formulate an appropriate hypothesis if your objective is to determine whether an MBA degree helps boost your salary **by more than \$45,000 on average**.

- What is the absolute value of the **test statistic?** (Round to 6 decimal places.) **2 . 035737**
- What is the **p-value?** (Round to 6 decimal places.) **0 . 036131**
- State your conclusions using $\alpha=5\%$: **# 1**

#1: Yes, we have sufficient evidence that having an MBA degree helps boost average salary by more than \$45,000 on average.

#2: No, we don't have sufficient evidence that having an MBA degree helps boost average salary by more than \$45,000 on average.

Answers:

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Response Feedback:

Question 7

10 out of 10 points



The data set **GM data.xlsx** contains retail prices and characteristics of used General Motors cars. Data were collected in 2005 from Kelly Blue Book.

VARIABLE DESCRIPTIONS:

Price: Suggested retail price of the used 2005 GM car.
Mileage: Number of miles the car has been driven.



Make: Manufacturer of the car such as Saturn, Pontiac, and Chevrolet.

Model: Specific models for each car manufacturer such as Ion, Vibe, Cavalier.

Trim (of car): Specific type of car model such as SE Sedan 4D, Quad Coupe 2D.

Type: Body type such as sedan, coupe, etc.

Cylinder: Number of cylinders in the engine.

Liter: A more specific measure of engine size.

Doors: Number of doors.

Cruise: Indicator variable representing whether the car has cruise control (1 = cruise).

Sound: Indicator variable representing whether the car has upgraded speakers (1 = upgraded).

Leather: Indicator variable representing whether the car has leather seats (1 = leather).

The following regression output was obtained in Excel for a multiple linear regression model, in which the dependent variable is **Price**:

Multiple Regression for Price		Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
Summary		0.6308	0.3979	0.3941	7694.076281	0	0
ANOVA Table		Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value	
Explained		5	31220732630	6244146526	105.4775687	< 0.0001	
Unexplained		798	47240650231	59198809.81			
Regression Table		Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
Constant		10366.3057	1953.596747	5.306266874	< 0.0001	6531.510175	14201.10123
Mileage		-0.093063172	0.067244239	-1.383957533	0.1668	-0.225059659	0.038933316
Liter		4027.65525	265.9595228	15.14386553	< 0.0001	3505.592347	4549.718153
Doors		-1093.994474	320.5992656	-3.412342419	0.0007	-1723.311978	-464.6769688
Cruise		7996.045657	1655.511696	4.829954192	< 0.0001	4746.373553	11245.71776
Mileage*Cruise		-0.104504509	0.07728864	-1.352132851	0.1767	-0.256217564	0.047208546

In this model, five variables including an interaction term were used to predict the selling price of a used GM car. Using $\alpha=5\%$, which of these variables have coefficients that are "**statistically significant**", meaning that these variables are good linear predictors of price? Select all that apply.

Selected Answers: Doors

Cruise

Liter

Answers:

Doors

Mileage * Cruise

Cruise

Liter

Mileage

Response Feedback: 😊



[See data Catalog](#)

Marketing.xlsx

The file contains data on 1,000 HyTex customers.

- **Age:** coded as 1 for 30 or younger, 2 for 31 to 55, 3 for 56 or older
- **Gender:** coded as 1 for males, 0 for females
- **OwnHome:** coded as 1 if the customer owns a home, 0 otherwise
- **Married:** coded as 1 if the customer is currently married, 0 otherwise
- **Close:** coded as 1 if the customer lives reasonably close to a shopping area that sells similar merchandise, 0 otherwise
- **Salary:** combined annual salary of the customer and spouse (if any)
- **Children:** number of children living with the customer
- **History:** coded as "NA" if the customer had no dealings with HyTex before this year, 1 if the customer was a low-spending customer last year, 2 if medium-spending, 3 if high-spending
- **Catalogs:** number of catalogs sent to the customer this year
- **FirstPurchase:** date of the customer's first purchase with HyTex
- **AmountSpent:** total amount of purchases made by the customer this year
- In addition, the variables **Region**, **State**, and **City** indicate where the customer resides.



HyTex wants to find some useful and quick information about its customers to strategize its marketing more effectively. How can it proceed? You are working in the Marketing Division of HyTex. You decide to estimate a linear regression equation for **Amount Spent**, using:

- **Age**,
- **Salary**, and
- **Region**: East, West, South, and Midwest.

You obtain the following regression output:

Multiple Regression for AmountSpent	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
Summary	0.7058	0.4981	0.495580	682.5801843	0	0
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value	
Explained	5	459621982	91924396.39	197.2983413	< 0.0001	
Unexplained	994	463120213.8	465915.708			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
Constant	-262.8213562	77.93126704	-3.372476366	0.0008	-415.7500458	-109.8926667
Age	132.0028871	33.71982442	3.914696752	< 0.0001	65.83267388	198.1731004
Salary	0.020816366	0.000764561	27.2265718	< 0.0001	0.019316028	0.022316705
D_West	86.18818744	62.0119335	1.38986454	0.1649	-35.50114324	207.8775181
D_South	64.3801093	60.92458607	1.056718042	0.2909	-55.17546143	183.93568
D_Midwest	84.56760166	60.52298954	1.397280642	0.1626	-34.19989473	203.3350981

- Do U.S. regions help with predicting the amount spent? (**YES or NO**) **[a]**
- Explain why: # **[b]**

#1: Because the coefficients make sense economically.

#2: Because the p-values are very high.

#3: Because the adjusted R-squared is high.

Selected Answer:



See data

Catalog Marketing.xlsx

The file contains data on 1,000 HyTex customers.

- **Age:** coded as 1 for 30 or younger, 2 for 31 to 55, 3 for 56 or older
- **Gender:** coded as 1 for males, 0 for females
- **OwnHome:** coded as 1 if the customer owns a home, 0 otherwise
- **Married:** coded as 1 if the customer is currently married, 0 otherwise
- **Close:** coded as 1 if the customer lives reasonably close to a shopping area that sells similar merchandise, 0 otherwise
- **Salary:** combined annual salary of the customer and spouse (if any)
- **Children:** number of children living with the customer
- **History:** coded as "NA" if the customer had no dealings with HyTex before this year, 1 if the customer was a low-spending customer last year, 2 if medium-spending, 3 if high-spending
- **Catalogs:** number of catalogs sent to the customer this year
- **FirstPurchase:** date of the customer's first



- purchase with
HyTex
- **AmountSpent:**
total amount of purchases made by the customer this year
 - In addition, the variables **Region, State, and City** indicate where the customer resides.

HyTex wants to find some useful and quick information about its customers to strategize its marketing more effectively. How can it proceed? You are working in the Marketing Division of HyTex. You decide to estimate a linear regression equation for **Amount Spent**, using:

- **Age**,
- **Salary**, and
- **Region**: East, West, South, and Midwest.

You obtain the following regression output:

Multiple Regression for AmountSpent	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
Summary	0.7058	0.4981	0.495580	682.5801843	0	0
ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value	
Explained	5	459621982	91924396.39	197.2983413	< 0.0001	
Unexplained	994	463120213.8	465915.708			
Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
Constant	-262.8213562	77.93126704	-3.372476366	0.0008	-415.7500458	-109.8926667
Age	132.0028871	33.71982442	3.914696752	< 0.0001	65.83267388	198.1731004
Salary	0.020816366	0.000764561	27.2265718	< 0.0001	0.019316028	0.022316705
D_West	86.18818744	62.0119335	1.38986454	0.1649	-35.50114324	207.8775181
D_South	64.3801093	60.92458607	1.056718042	0.2909	-55.17546143	183.93568
D_Midwest	84.56760166	60.52298954	1.397280642	0.1626	-34.19989473	203.3350981

- Do U.S. regions help with predicting the amount spent? (**YES or NO**) **NO**
- Explain why: **# 2**

#1: Because the coefficients make sense economically.

#2: Because the p-values are very high.

#3: Because the adjusted R-squared is high.

Answers:

- 1
- 2
- 3
- YES
- NO

Response Feedback:

Question 9

10 out of 10 points



You have performed hypothesis testing for the population mean and obtained the following results for the one-tailed test:

Test statistic = 1.9317

P-value = 0.0612

Using $\alpha = 5\%$, of course, you don't reject the null hypothesis. This is disappointing because it means that your research hypothesis is not supported by your data.

What can you do about your data (improper data manipulation is not allowed!) or the methodology in order to achieve the opposite conclusion? **Click on all valid options from the list below.**

Selected Consider removing outliers
Answers:

Collect more data

Critically re-evaluate your sample: if it's not a random sample, then consider collecting a different sample

Increase level α

Answers: Divide the p-value by 2

Decrease level α

Increase the sample mean

Change the test from one-tailed to two-tailed

Consider removing outliers

Collect more data

Take out some observations from the data randomly, in order to reduce the sample size

Critically re-evaluate your sample: if it's not a random sample, then consider collecting a different sample

Reduce the sample standard deviation

Buy yourself a pumpkin spice latte from Starbucks and do nothing because there's absolutely nothing you can do to change the results and conclusions

Increase level α

Multiply the test statistic value by 2

Response Feedback:  **Bingo!**

Saturday, December 1, 2018 10:32:50 PM EST

← OK