Problem 1 - West Campus Rent

Let's use data to explore a question about the Austin real-estate market: to what extent do West Campus apartment rents decrease as you get further away from the center of campus? The file apartments.csv contains a data frame with a sample of 67 apartment buildings in the West Campus area.

Build a linear regression model to predict the rent of a two-bedroom apartment in West Campus from the variables listed above: sqft, miles_from_tower, furnished, pool, laundry, electricity, and water. Use this model to address two questions:

I built a linear model that
I first began by creating a linear regression model to predict the variation of rent of apartments in west campus depending on the variables: How many square feet, how many miles the apartment is from the UT tower, if the apartment is furnished, if the apartment has a pool or not, laundry, electricity, and water.

Part A

What is the drop-off in price associated with an **extra 0.1-mile walking distance** to the UT Tower, adjusting for the other apartment features in the model (including size of unit)? (Pay attention to the units of the variables here.)

Part B

What is the increase in rent associated with an **additional 100 square feet in size**, adjusting for the other features of the apartment in the model (including distance from the UT Tower)? (Again, pay attention to variable units in your interpretation.)

Part C

Include in the Results section of your write-up two scatter plots to (1) visualize the *overall* relationship between rent and miles from campus as well as (2) the *overall* relationship between rent and apartment size. Include appropriate caption(s) identifying key elements of the plots and specifying variable units. You might consider using a jitter plot for rent versus miles_from_tower to differentiate overlapping observations in your display.

Question:

1a)What is the variation/drop-off in prices in apartments when the distance from the apartment is an extra 0.1 mile walking distance from the UT tower?

1b)How does rent for apartments in west campus increase with the addition of 100 square feet in size?

1c)What is the relationship between west campus apartment rents and how many miles the apartment is from the UT tower?What is the overall relationship between west campus apartment rent and apartment size?

Approach:

1a)I first began by using an unadjusted linear regression model to account for the monthly rent of a two-bedroom apartment in west campus by the amount of miles the apartment is from the UT tower. I then created an adjusted linear regression model to account for variation of rent of apartments in west campus depending on the variables: How many square feet, how many miles the apartment is from the UT tower, if the apartment is furnished, if the apartment has a pool or not, laundry, electricity, and water. Finally, I bootstaped the data and concluded the confidence intervals.

1b)I built a linear regression model to represent the miles for the west campus apartment to the UT tower and how that affects the tent of west campus apartments. I then created an adjusted model to account for an additional 100 square feet and the variation of square feet, if the apartment is furnished, if the apartment has a pool or not, laundry, and if electricity and water is included. I bootstrapped the data and concluded the confidence intervals. I then created a scatter plot to display the variation of square feet and its impact on rent.

1c)I created a scatter plot to show the overall relationship between how the amount of miles the west campus apartment is from the UT tower and how that relationship affects rent. I then created another scatter plot to show the overall relationship between the sizes of a west campus apartment and how that affects rent.

Result:

1a) As a result of the unadjusted linear regression model I got the following results:

	name	lower	upper	level	method
1	Intercept	1227.231	1506.927	0.95	percentile
2	miles_from_tower	-836.881	-421.774	0.95	percentile
3	sigma	112.893	214.372	0.95	percentile
4	r.squared	0.090	0.449	0.95	percentile
5	F	6.447	52.888	0.95	percentile

For my unadjusted model I ended up with the equation:

Unadjusted model:

Rent = **1366.8075 + -63.5061* miles_from_tower**

(1366_ represents our bass line rent and (-63) represents the coefficient, which also represents how much rent decreases based on every 0.1 mile closer a west campus apartment is from the UT tower.

Adjusted Model:

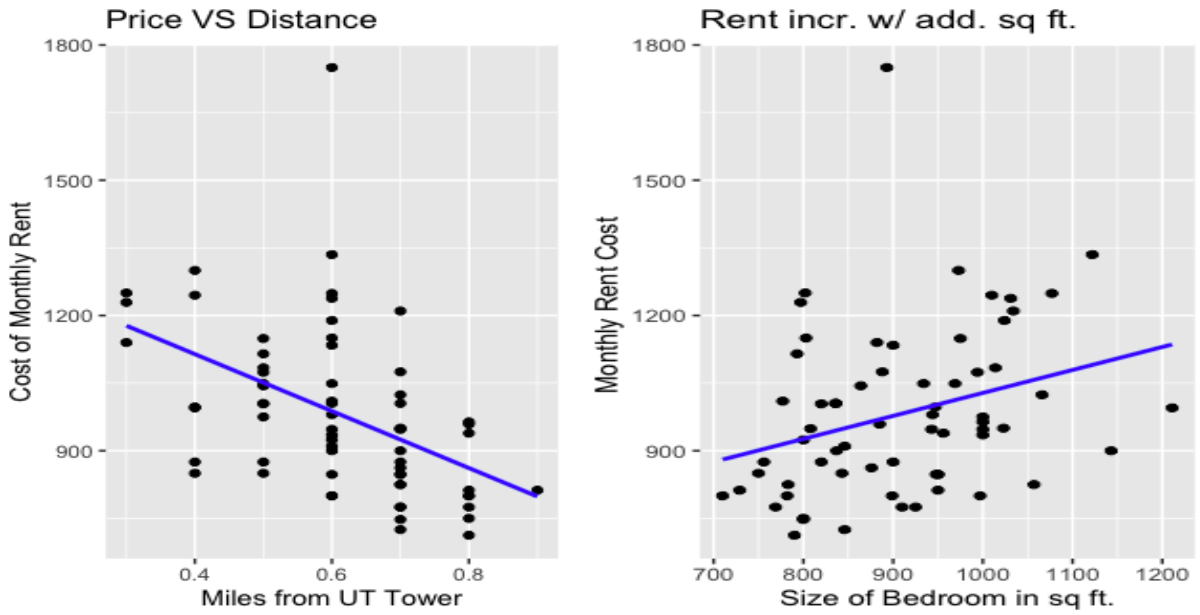
Rent = **795 – 38*distance + 38* miles_from_tower + 163*furniture -48*pool + 104*laundry_in_unit + 64*electricity – 23*water**

1b)The result of a linear model that showed the relationship between rent of a west campus apartment based on how many miles it is from the UT tower. The adjusted linear model that accounts for the other variables resulted in the following equation:

Adjusted Model:

Rent = 795 – 38*distance + 38*sf + 163*furniture -48*pool + 104*laundry_in_unit + 64*electricity – 23*water

1c) As a result of scatterplots is:



Conclusion:

1a)From the derived linear equation and our confidence interval, we can conclude with 95% confidence that for each 0.1 extra miles a west campus apartment is to the UT tower, we can expect rent to decrease by a factor of $39 for the bases lines rent of $795 when the other variables such as, how many square feet, how many miles the apartment is from the UT tower, if the apartment is furnished, if the apartment has a pool or not, laundry, electricity, and water are not accounted for. We get the confidence intervals of (0.063, 0.695. 63 represents the slope and the drop off price per extra 0.1 mile. We can conclude that the drop off price for a west campus apartment when the distance from the apartment is an extra 0.1 mile walking distance from the UT tower is between $68 to $7 per 0.1 mile. When we account for the other variables, we can expect that rent decreases by a factor of $63 from the base rent of $1363 for every 0.1 miles the apartment is from the UT tower.

1b) From my linear equations we can conclude that On average for every additional 100 square foot rent for West Campus Apartments increase (0.063, 0.695)between $6.3 and $69.5 onv avg for every add. 100 sq. ft.

Problem 2 - P-hacking with Green Buildings

Background: In this problem, you'll see how "p-hacking" actually works! Remember, p-hacking is NOT a recommended practice—quite the opposite. It's something to be avoided and for which to be on the lookout in evaluating others' work. Thus, the point of Problem 2 is to sensitize you to the range of possible choices that one can make in a data analysis; it's this sheer range of choices that makes p-hacking even possible.

Question:

2a)The question I am solving is what is the most convincing analyst in which I can find thay the green certification is a "statistically significant" predictor of success on the commercial real estate market?

2b)The question I am solving is what is not a convincing analyst in which I can find thay the green certification is a "statistically significant" predictor of success on the commercial real estate market?

2c)The question we are trying to solve is, What is more plausible, that green certification is a statically significant predictor of success on the commercial real estate market, or that green certification does not have a statically significant predictor of success on the commercial real estate market?

Approach:

2a)To run a convincing anylist that proves that green certifications is a significant predictor of success on the commercial real estate market I first began by creating a linear model to represent the increase of the commercial real estate market success due to having a green certification. I used the variable rent.

2b)To run a convincing anylist that proves that green certifications is not a significant predictor of of success on the commercial real estate market I first began by creating linear model between green_rating (green certifications) and Rent(represents significant predictor of of success)

Result:

2a) As a result of my linear model my intercepts intercept -5.38 and -3.43.

2b)As a result of linear model my intercepts were -1.74 and -0.18 .

Conclusion:

2a)Based on the p-hacking data we can conclude that the 95% confidence intervals suggest that green certification is a "statistically significant" predictor of success on the commercial real estate market, because out p value is so small that signifies some positive relationship between green certification and the success on the commercial real estate market.

2b)Based on the data collected the P value proves that there is not a significant impact on green certification on Commercial Real Estate. The intercepts goes through 0 there for not having a strong relationship between green certification and Commercial Real Estate.

Part C

	Part A	Part B	Which is Better
Outcome Choices	Success based on square foot	Success is based on revenue per square foot	For A & B we should calculate success by measuring the revenue rather than measuring the total rent.

Green Certification calculation	Green_ratings both the LEED and energy star	LEED	A: Green_ratings is the best selection because it includes all of the green certificates
Approach	Matchit and confidence intervals	Matchit and confidence intervals	A&B: Match it would be the more useful in a line of linear regression to find the confidence interval
Adjusted cofounder	Market rent, age, class, city	Market rent, age, class, city	Rent is the most useful in accounting and adjusted variables

Problem 3 - Hotel ML

The goal of this problem is simple: use linear regression to build a Machine Learning model to predict whether the guest party associated with a hotel booking will include any children.

Question:

3)The question we are trying to solve is whether or not a guest party at one of the mentioned hotels will bring children?

Approach:

To solve this question I created four different linear models to outline. My first linear model also known as my small model is created by only using market segments, adults, customer types, and is the guest of repeat guests variables. My next model I created was titled my big model. the big model attributed it for

all possible predictors except the arrival date variable. The next model I created title the huge model attributed for all possible predictors except the arrival date variable. My fourth and final model accounted for all possible predictors and the additional month of the year based on the arrival date variable.

Result:

Model	In-Sample RMSE	Out-of-Sample RMSE
Small model	.3475	.3477
Big model	.3190	.3197
Huge model	.3102	.5205
Big model pt.2	.2956	.2940

Conclusion: According to the data in the linear model created and the models described above We can say with 95% certainty that the likelihood of bring a chid is between .2956 and .2940.

Problem 4 - Monte Carlo Investment simulation

The graph will not load however this is the code to retrieve the graf:

```
ggplot() +
  geom_line(aes(x = 1:horizontal, y = wealth_time)) +
  +labs(
    title = "Line of Wealth 10 risk",
    x = "horizontal"
    y = "we**alth" )**
```

4a)The graph above shows how the wealth fluctuates as you bet 10000 rounds with a 10% of your initial wealth. based on the graphs it would appear that your wealth will decrease leaving you with virtually nothing.

My graft will not load onto my Google doc however here is the code that I use to derive my graft:

```
ggplot() +

  geom_line(aes(x = 1:horizontal, y = wealth_time)) +

  +labs(

    title = "Line of Wealth .5% risk",

    x = "horizontal"

    y = "wealth"  )
```

4b)The graphs indicate that That the wealth will fluctuate and eventually gain you a return investment.

My graft will not load onto my Google doc however here is the code that I use to derive my graft:

```
ggplot() +

  geom_line(aes(x = 1:horizontal, y = wealth_time)) +

  +labs(

    title = "Line of Wealth 5% risk",

    x = "horizontal"

    y = "wealth"  )
```

4c) The best value of C to return investment is a 5% risk.I derived this conclusion by calculating how much return will be lost in a variety of .5% risk up to 10% risk the risk 10% will leave you broke and a risk of .5 % will leave you basically with the same amount you began with with little variety. However a 5% risk will gain you Revenue but has the possibility also lose Revenue however it's the best chance I gaining wealth.