Lab 15

Stat 427

Fall 2020

Instructions

Complete all questions. To prepare for the randomly collected lab, follow the instructions on the class website to prepare the work for submission. These submission rules will apply to all labs throughout the semester.

Introductory inferential methods

(1) *It ain't easy bein' green.* A dealer in recycled paper places empty trailers at various sites. The trailers are gradually filled by individuals who bring in old newspapers and magazines, and are picked up on several schedules. One such schedule involves pickup every second week. This schedule is desirable if the average amount of recycled paper is more than 1600 cubic feet per 2-week period. The dealer's records for eighteen 2-week periods show the following volumes (in cubic feet) at a particular site: the data is given in the `R` code section at the end of this document.
   (a) Estimate the true mean weight of recycled paper with 95% confidence. Interpret.
   (b) Is there sufficient evidence that the mean amount of recycled paper is more than 1600 cubic feet per 2 week period? Conduct a hypothesis test.

(2) *Trees* A study of 584 longleaf pine trees in the Wade Tract in Thomas County, Georgia had several purposes. To see if there is a difference in their sizes (in diameters) in two separate areas of the Wade Tract (northern and southern areas), a random sample of 30 trees from the northern area and 30 trees from the southern area was taken. the data is given in the `R` code section at the end of this document.
   (a) Estimate the true difference in mean tree sizes between the northern and southern parts of the Wade Tract with 95% confidence
   (b) Is there a significant difference in the mean diameter of trees in the north versus the trees in the south? Conduct hypothesis test

(3) Breakfast is the most important meal of the day, or so the experts have said for decades. In one office, there was a bet to see if the hours worked was greater for those that ate breakfast daily than those that did not eat a daily breakfast. The data is given in the `R` code section at the end of this document.
   (a) Conduct the hypothesis test to see if the mean difference of worked hours is greater with breakfast eaten than not eating breakfast
   (b) Estimate the true mean difference with 95% confidence

(4) A six-sided die is rolled 120 times. The data in the following table are the result of the 120 rolls. Conduct a hypothesis test to determine if the die is fair (does the data follow a uniform distribution, i.e. every side has an equal probability)

Face value	Frequency
1	15
2	29
3	16
4	15
5	30
6	15

(5) A major food manufacturer is concerned that the sales for its skinny french fries have been decreasing. As a part of a feasibility study, the company conducts research into the types of fries sold across the country to determine if the type of fries sold is independent of the area of the country. The results of the study are shown in the following table. Is there evidence that fry preference is independent of location? Conduct a hypothesis test.

Type of Fries	*Northeast*	*South*	*Central*	*West*	**Total**
skinny	70	50	20	25	165
curly	100	60	15	30	205
steak	20	40	10	10	80
Total	190	150	45	65	450

(6) A video game developer is testing a new game on three different groups. Each group represents a different target market for the game. The developer collects scores from a random sample from each group. Is there evidence that the mean scores are the same for all target groups or is at one target group mean score different? Conduct hypothesis (anova) test. If appropriate, conduct Tukey's HSD.

(7) *Carbonation*: Corrosion of steel reinforcing bars is the most important durability problem for reinforcing structures. Carbonation of concrete results form a chemical reaction that lowers the pH value by enough to initiate corrosion of the rebar. Data on the carbonation depth (mm) and strength (MPa) for a sample of core specimens was taken from a particular building, and all the regression output is provided. We are interested in modeling the strength[1]
(a) State the regression model and define its components
(b) Looking at the raw data scatterplot, does it appear as if there is a linear relationship? Positive or negative slope?
(c) State the equation of the regression equation (from output). Use it to estimate the strength when the carbonation depth is 8 mm and estimate it again when the depth is 20 mm
(d) Calculate the residuals for both of your estimates in part c. The observed value for 8 mm is 22.8 MPa $((8, 22.8))$ and for 20 mm is 17.1 MPa $((20, 17.1))$
(e) Do a significance test of the slope. State hypotheses, t statistic, $pvalue$, results, and conclusion of the test
(f) State, define, and describe R^2 and r (R^2 is on the output and r will require a calculation from the output)
(g) List assumptions of regression. Are the assumptions of regression met? Briefly explain how each assumption is met or not
(h) How is the model? Good, bad, etc.? Give specific evidence (use answers from parts e, f, and g)

[1] "The Carbonation of Conrete Structures in the Tropical Environment of Singapore" (*Magazine of Concrete Research*, 1996: 293-300)

Data for some problems

Problem 1

```
recycle=c(1935,1556,1752,1969,1804,1842,1994,1810,1827,1725,2003,1499,1809,1795,1622,
          1620,1777,2035)
```

Problem 2

```
diameter=c(27.8,14.5,39.1,3.2,58.8,55.5,25,5.4,19,30.6,15.1,
        3.6,28.4,15,2.2,14.2,44.2,25.7,11.2,46.8,36.9,54.1,
        10.2,2.5,13.8,43.5,13.8,39.7,6.4,4.8,44.4,26.1,50.4,
        23.3,39.5,51,48.1,47.2,40.3,37.4,36.8,21.7,35.7,32,
        40.4,12.8,5.6,44.3,52.9,38,2.6,44.6,45.5,29.1,18.7,
        7,43.8,28.3,36.9,51.6)
direction=c(rep('North',each=30),rep('South',each=30))
trees=data.frame(diameter,direction)
```

Problem 3

```
w=c(8,7,9,5,9,8,10,7,6,9)
wo=c(6,5,5,4,7,7,7,5,6,5)
hours=c(w,wo)
eat=c(rep('With',each=length(w)),rep('Without',each=length(w)))
breakfast=data.frame(hours,eat)
```

Problem 6

```
scores=c(101,108,98,107,111,151,149,160,112,126,101,109,198,186,160)
group=c(rep('A',each=5),rep('B',each=5),rep('C',each=5))
videogame=data.frame(scores,group)
```

Problem 7

```
carbonation=c(8,15,16.5,20,20,27.5,30,30,35,38,40,45,50,50,55,55,59,65)
strength=c(22.8,27.2,23.7,17.1,21.5,18.6,16.1,23.4,13.4,19.5,12.4,13.2,11.4,10.3,14.1,9.7,12,6.8)
```