Review problems

Modules 7-12

Stat 2510

Problems

(1) A company that manufactures coffee for use in commercial machines monitors the caffeine content in its coffee. The company randomly selected 50 samples of coffee every hour from its production line and determines the caffeine content, which should usually be 107 mg. From historical data, the standard deviation is 7.1 mg and caffeine content is known to to be approximately normal. From one hourly sample of 50, the mean caffeine content is 110 mg.
   (a) Is there sufficient evidence that the mean caffeine content is more than the usual amount?
   (b) State the kind of error that could have been made and describe it within the context of the data
   (c) Estimate μ, the true average caffeine content of the coffee, with 95% confidence. Interpret

(2) It is thought that more than 70% of all faults in transmission lines are caused by lightning. In a random sample of 200 faults from a large data base, 151 are due to lightning.
   (a) Is there sufficient evidence that the proportion of faults in transmission due to lightning strikes is different from 70%? Conduct a hypothesis test
   (b) State the kind of error that could have been made and describe it within the context of the data
   (c) Estimate p, the true proportion of faults in transmission due to lightning strikes, with 95% confidence. Interpret

(3) In 1882 Michelson measured the speed of light (usually denoted as c in Einstein's equation $E = mc^2$). He reported the results of 23 random trials with a mean of 29756.22 km/sec and standard deviation of 107.12 km/sec. [Note that the actual speed of light is actually 299,792,458 metres per second but remember, these results are from experiments done in *1882*; use the units of measurement that Michelson used (km/sec).]
   (a) Suppose previous experiments of Michelson found that the speed of light was 29750 km/sec. Is there sufficient evidence from his experiments that the speed of light is significantly different from the previous result of 29750? Let $\alpha = 0.02$.
   (b) State the kind of error that could have been made and describe it within the context of the data
   (c) Estimate μ, the true speed of light with 98% confidence. Interpret

(4) Some Final Fantasy games have jobs for characters to utilize. A job in the game means that characters are constrained to the weapons and armor they can use (as an example, a mage cannot wear plated armor and cannot use swords). As a player, I would like to compare the average strength stat of Knights and Thieves at level 50. Below are the summary statistics for 10 strength stats for knights and 10 strength stats for thieves, all taken at level 50.

(a) Is there sufficient evidence that the strength stats for knights are significantly higher than the strength stats for thieves at level 50?

(b) Estimate the true difference in means of the strength stats of knights and thieves with 95% confidence and interpret

(c) State the kind of error that could have been made and describe it within the context of the data

```
==================================
Statistic N   Mean  St. Dev. Min Max
----------------------------------
knight    10 76.700  2.830   72  81
thief     10 70.000  2.944   65  75
----------------------------------
```

(5) *Paired t-test and CI (dependent samples)* μ_d: Let's say you're playing Final Fantasy XIV and you want to test if a specific ability (Berserk) increases your damage output. You measure your damage output (in terms of DPS, or damage per second) in two different situations: before using Berserk and after using Berserk. Data was collected from 10 different runs where the same battle was performed, once without Berserk and once with Berserk.

(a) Is there sufficient evidence that there is a significant difference in damage output by not using or using Berserk

(b) Estimate the true mean difference in DPS before and after the using Berserk with 95% confidence and interpret

(c) State the kind of error that could have been made and describe it within the context of the data

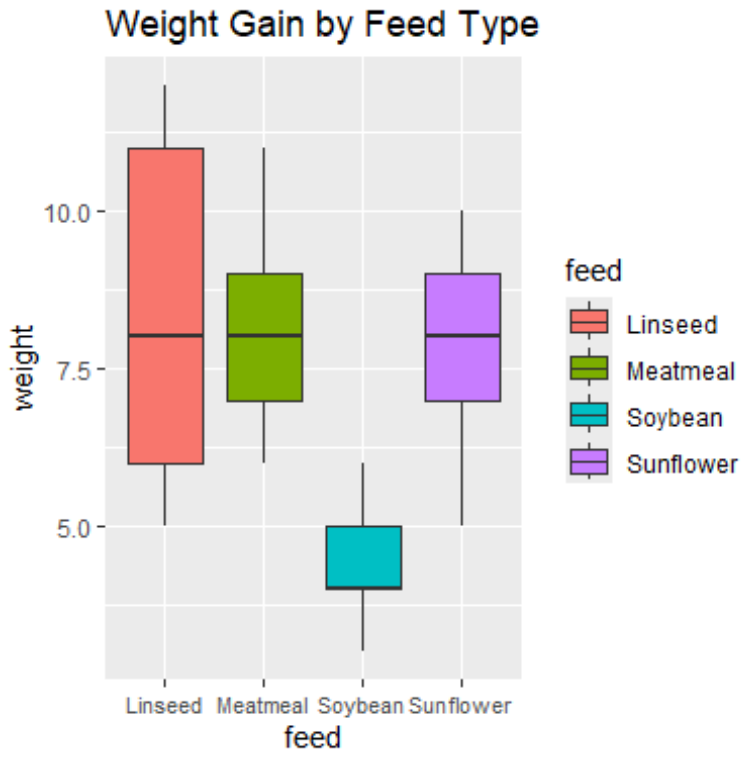
```
   run before after diff
1    1   1500  1800  300
2    2   1600  1900  300
3    3   1550  2000  450
4    4   1700  2100  400
5    5   1450  1750  300
6    6   1600  1850  250
7    7   1620  1930  310
8    8   1580  1900  320
9    9   1540  1800  260
10  10   1650  2000  350
```

```
====================================
Statistic N   Mean   St. Dev. Min Max
------------------------------------
diff      10 324.000  61.319  250 450
------------------------------------
```

(6) *Difference of two proportions test and CI*: A hospital administrator suspects that the delinquency rate in the payment of hospital bills has increased over the past year. Hospital records show that the bills of 48 of 1284 persons admitted in the month of April have been delinquent for more than 90 days. This number compares to 34 of 1002 persons admitted during the same month one year ago.
   (a) Is there sufficient evidence that there is an increase in delinquency rate in the payment of hospital bills over the last year?
   (b) Estimate the true difference of proportions of delinquent bills over the last year with 95% confidence and interpret
   (c) State the kind of error that could have been made and describe it within the context of the data

(7) analyze data from an experiment where treatments are assigned to experimental units entirely by chance. A classic example involves comparing the effects of different treatments on a specific outcome, such as the effect of various feeds on chicken weight gain. This randomized experiment is to access the weight gain (in grams) of 20 chicks fed one of four different feed types (A, B, C, and D). Five chicks were randomly assigned to each feed type.
   (a) State hypotheses
   (b) Calculate ANOVA table
   (c) Find critical value of F
   (d) Conclude the test (results and conclusion)
   (e) List assumptions
   (f) Use provided output to conduct a multiple comparison using Fisher's LSD test

```
           ni ybari s2i
Linseed    5   8.4 9.3
Meatmeal   5   8.2 3.7
Soybean    5   4.4 1.3
Sunflower  5   7.8 3.7
```




```
Study: fit ~ "feed"

LSD t Test for weight

Mean Square Error:  4.5

feed,  means and individual ( 95 %) CI

           weight      std r        se      LCL       UCL Min Max Q25 Q50 Q75
Linseed       8.4 3.049590 5 0.9486833 6.388881 10.411119   5  12   6   8  11
Meatmeal      8.2 1.923538 5 0.9486833 6.188881 10.211119   6  11   7   8   9
Soybean       4.4 1.140175 5 0.9486833 2.388881  6.411119   3   6   4   4   5
Sunflower     7.8 1.923538 5 0.9486833 5.788881  9.811119   5  10   7   8   9

Alpha: 0.05 ; DF Error: 16
Critical Value of t: 2.119905

least Significant Difference: 2.844151

Treatments with the same letter are not significantly different.

           weight groups
Linseed       8.4      a
Meatmeal      8.2      a
```

```
Sunflower    7.8       a
Soybean      4.4       b
```

(8) *SLR (simple linear regression)*: Given is a dataset for a survey that asked respondents their years of experience (`yrsexp`) at a job and their salary (`salary`) amount. Of interest is to model salary by years of experience

(a) State the population model equation 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 regression equation using the provided output. Using the regression equation, estimate salary when years of experience is 1.1 and estimate it again when the years of experience is 9.6.

(d) Calculate the residuals for one of your estimates in part c. The observed value of salary for 1.1 years of experience is $39,343 ($(1.1,39343)$) and for 9.6 years of experience is $112,635.00 ($(9.6,112635)$)

(e) Interpret slope and intercept *in context* of the data. If something does not make sense in context, state it and describe why

(f) What is the coefficient of determination, R^2? List the value and interpret in context

(g) What is the correlation, r? List the value and interpret in context

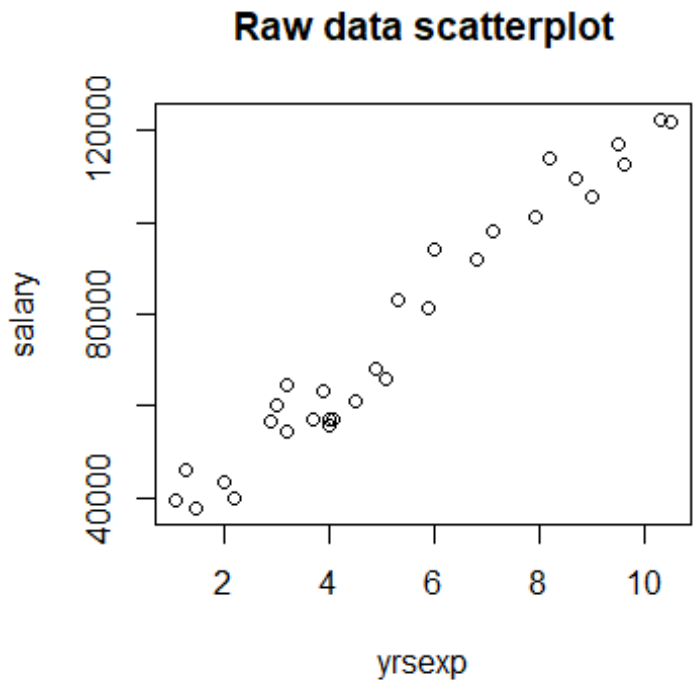
(h) Is the slope significant? Conduct hypothesis test; include hypotheses, test statistic, results, and conclusion

(i) List assumptions of regression (words or symbols)

(j) Using parts f, g, and h, is this a good model? Reference those to verify your claim (briefly describe)

	yrsexp	salary
1	1.1	39343
2	1.3	46205
3	1.5	37731
4	10.3	122391
5	10.5	121872
6	2.0	43525
7	2.2	39891
8	2.9	56642
9	3.0	60150
10	3.2	54445
11	3.2	64445
12	3.7	57189
13	3.9	63218
14	4.0	55794
15	4.0	56957
16	4.1	57081
17	4.5	61111
18	4.9	67938
19	5.1	66029
20	5.3	83088
21	5.9	81363

```
22    6.0  93940
23    6.8  91738
24    7.1  98273
25    7.9 101302
26    8.2 113812
27    8.7 109431
28    9.0 105582
29    9.5 116969
30    9.6 112635
```




```
Call:
lm(formula = salary ~ yrsexp, data = expsal)

Residuals:
    Min      1Q  Median      3Q     Max
-7958.0 -4088.5  -459.9  3372.6 11448.0

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  25792.2     2273.1   11.35 5.51e-12 ***
yrsexp        9450.0      378.8   24.95  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5788 on 28 degrees of freedom
```

```
Multiple R-squared:  0.957, Adjusted R-squared:  0.9554
F-statistic: 622.5 on 1 and 28 DF,  p-value: < 2.2e-16
```