

## Stat 225 - Review for exam 3

The second exam will be this coming Monday, November 19. Here are some problems that might help. Note that you may use a simple calculator.

1. A random sample of 18 college texts compared their local bookstore price to their amazon.com price. The data indicate that, on average, the local price minus the amazon price was \$22 with standard deviation of \$4. We are curious if the amazon price is generally *less* than the book store price.
  - (a) Write down the hypothesis test.
  - (b) Compute the standard error for the problem and test statistic.
  - (c) State what test you plan to use for this problem and why
  - (d) What is the conclusion of the test and what is your supporting evidence?
2. A recent study surveyed people's attitudes on the relative importance of economic issues. In particular, the study examined the question - what is more important:
  - Reducing the federal deficit or
  - Cutting taxes for corporations

The results are summarized in table 1. Let  $p_D$  denote the proportion of democrats who

	Democrats	Republicans
Deficit over taxes	97	66
Taxes over deficit	43	48

Table 1: Views on deficit reduction over tax reduction

think that reducing the deficit is the more important of the two issues and let  $p_R$  denote the proportion of republicans who think that reducing the deficit is the more important of the two issues. We suspect that  $p_D < p_R$  so let's use a hypothesis test to examine this question.

- (a) Compute the observed proportions  $\hat{p}_R$  and  $\hat{p}_D$ , as well as the difference

$$\hat{p} = \hat{p}_D - \hat{p}_R.$$

- (b) Compute the standard error and the test statistic.
- (c) Use a normal table to compute the  $p$ -value.
- (d) State the conclusion of the test.
- (e) Why was it OK to use a normal table?

3. A statistics professor teaching two sections of the same introductory statistics course is concerned that exam scores in one section (let's call it section A) are noticeably lower than exam scores in another section (section B). To test this hypothesis, he computed the mean and standard deviation for the combined scores on the first two exams for both classes and found the following:

**Section A:** Mean=118.4, std dev = 40, number of students = 35

**Section B:** Mean=133.5, std dev = 32.7, number of students = 32

Viewing these data as samples from a random process, find

- The difference between the two means.
  - The associated standard error and test statistic.
  - Compute the  $p$ -value using a normal distribution.
  - From the computations, can we conclude with a 99% level of confidence that section A scores lower than class B?
4. A statistics professor teaching two sections of the same calc based statistics course wants to compare exam scores in one section (let's call it section A) to another section (section B). To test if they are the same, he computed the mean and standard deviation for the combined scores on the first two exams for both classes and found the following:

**Section A:** Mean=112, std dev = 30, number of students = 14

**Section B:** Mean=123.5, std dev = 18.7, number of students = 12

Viewing these data as samples from a random process, find

- The difference between the two means.
  - The associated standard error and test statistic.
  - Find the critical  $t^*$  value for a 95% level of confidence.
  - From the computations, can we conclude with a 95% level of confidence that section A scores lower than class B?
5. Find the regression line for the data  $\{(0,0), (1,0), (2,2)\}$ .
6. Suppose that I have paired data  $X$  and  $Y$  that are well correlated with  $r = 0.77$ ,  $\bar{x} = 12$ ,  $\bar{y} = 25$ ,  $\sigma_x = 2$ , and  $\sigma_y = 4$ . Find the regression line for this data.
7. Suppose that  $X$  and  $Y$  have the following joint distribution table:

$X \backslash Y$	0	2	5	Total
0	0.1	0.2	0.2	
1	0.3	0.1	0.1	
Total				

Find the marginal distributions,  $E(X)$ ,  $E(Y)$ ,  $E(XY)$  and use this to check  $X$  and  $Y$  for independence.

8. Let  $f_c(x, y) = c(x^2 + y^3)$ .
- Find  $c$  so that  $f_c$  is a good probability distribution function on  $[0, 1] \times [0, 1]$ .
  - Using  $f = f_c$  where  $c$  is your answer above, find  $f_X$ , the marginal distribution of  $X$ .

9. Using the following Python commands:

```
from scipy.stats import linregress
import pandas as pd
df = pd.read_csv("https://www.marksmath.org/data/cdc.csv")
sample = df.sample(35, random_state=5)
linregress(sample.height, sample.weight)
```

I obtain the following result:

```
LinregressResult(slope=5.36, intercept=-187.27, rvalue=0.452,
                 pvalue=0.00632, stderr=1.838)
```

What is the purpose of this test and what conclusions can you draw?