Math 214 Statistics Fall 13 Midterm 3a

Name: Solutions

- Do any 4 of the following 5 questions.
- You may use a calculator, but no notes.

1	10	
2	10	
3	10	
4	10	
5	10	
	40	

Midterm 3	
Overall	

1. A study was conducted to determine if there was a difference in the driving ability of students from West University and East University by taking random samples of the students. Of the 100 sampled from West University, 15 had been involved in a car accident within the past year. Of the 120 randomly sampled students from East University, 14 students had been involved in a car accident within the past year.

   (a) Find a 90% confidence interval for the proportion of students at East university who have been in a car accident in the past year.

   (b) What is the pooled estimate of the proprtion of students who have been in accidents at both universities?

   (c) Bob wishes to test if the two universities' students have different accident rates. He sets his null hypothesis as $H_0: p_1 \neq p_2$. Explain his error.

a) $\hat{p}_2 = \frac{14}{120}$ $n = 120$ $z_* = 1.645$ $\hat{p}_2 \pm z_* \sqrt{\hat{p}_2(1-\hat{p}_2)/n}$

$= 0.1167$

~~(0.0874, 0.14~~

(0.0685, 0.1649)

b) $\hat{p} = \frac{15+14}{100+120} = 0.1318$

c) H_0 must involve definite values for the paramates, e.g. $p_1 = p_2$.

2. Using the data from the previous question test whether the two populations are significantly different at the 5% significance level.

   (a) State the null hypothesis and the alternate hypothesis.

   (b) Specify the test statistic and its distribution.

   (c) Calculate the test statistic, and state the critical value.

   (d) Would you reject or fail to reject H_0 at the given significance level? State your conclusion.

a) $H_0 : p_1 = p_2$

$H_a : p_1 \neq p_2$

b) $$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$$

c) $\hat{p}_1 = \frac{15}{100}$ $\hat{p}_2 = \frac{14}{120}$ $n_1 = 100$ $n_2 = 120$ $\hat{p} = \frac{15+14}{100+120} = 0.1318$

test statistic is 0.7278

critical value: $z_\alpha = 1.960$

d) fail to reject H_0: no significant evidence that proportions are different.

3. A quality control manager has concerns about the occurrence of defects on a component assembly used in the fabrication of an aircraft wing. The manufacturing plant operates with three shifts, and there were three particular types of defects that were of concern. The following table shows the count of defects of each type across the shifts from a random sample of 300 assemblies with defects from a week of production:

	Defect Type			
Shift	A	B	C	Total
Morning	43	36	56	135
Evening	17	49	29	95
Overnight	14	42	14	70
Total	74	127	99	300

(a) What is the appropriate null hypothesis in this situation? What test should you use?

(b) Assuming H_0, what is the expected number of type B defects in the evening shift?

(c) What is the contribution to the chi-square stastic for the type B defects in the evening shift?

(d) What are the degrees of freedom df for the chi-square statistic?

(e) Suppose the chi-square statistic for this data is 26.8. What is the conclusion? What does it mean in this case?

a) H_0: no association between rows and columns
H_a: some association.
Chi-square test.

b) expected $= \dfrac{(\text{row total}) \times (\text{col total})}{\text{total}} = \dfrac{95 \times 127}{300} \approx 40$

c) χ^2 contribution: $\dfrac{(\text{expected} - \text{observed})^2}{\text{expected}} = \dfrac{(49-42)^2}{40} \approx 1.926$

d) df $= (\text{rows}-1) \times (\text{cols}-1) = 4$

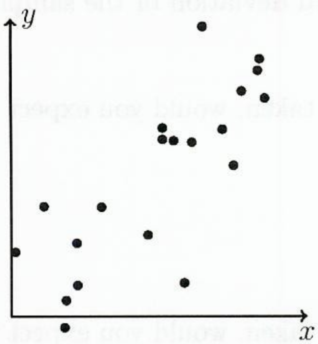
e) critical value for χ^2 df=4 at say 5% significance level is 11.14

reject H_0: some association between shifts and defects.

(p-value is ~~0.975~~ < 0.1%.)

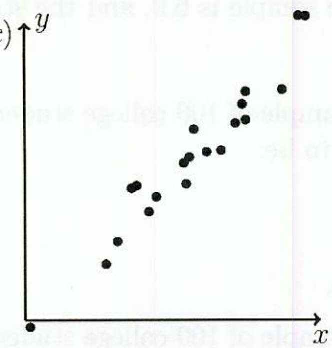
4. Guess the correlation coefficients for the following data.

b)



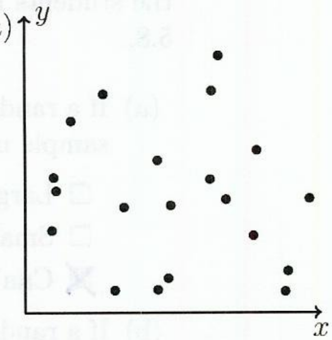
0.4

c)



0.9

a)



0

5. A random sample of 12 college students is taken, and the students are asked how many pairs of shoes they own. The average number of pairs of shoes of the students in the sample is 6.9, and the standard deviation of the sample is 5.8.

(a) If a random sample of 100 college students is taken, would you expect the sample mean to be:

- ☐ Larger.
- ☐ Smaller.
- ☒ Can't say.

(b) If a random sample of 100 college students is taken, would you expect the standard deviation to be larger or smaller?

- ☐ Larger.
- ☐ Smaller.
- ☒ Can't say.

(c) What can you say about the confidence interval for the population mean of the sample of size 10 compared with the sample of size 100?

- ☒ Larger.
- ☐ Smaller.
- ☐ Can't say.

(d) The students are also asked about how many hats they own, and the correlation coefficient for number of hats and number of shoes per student is $r = 0.85$. This means students with more shoes tend to have

- ☒ more hats.
- ☐ fewer hats.
- ☐ Can't say.

(e) Which of the following statements is correct, if any.

- ☐ Having more shoes causes students to have more hats.
- ☐ Having more hats causes students to have more shoes.
- ☒ Students who have more shoes tend to have more hats.

Math 214 Statistics Fall 13 Midterm 3a Lab Solutions

1. (40 points) A population of jellyfish have weight normally distributed with mean 0.9kg and standard deviation 0.5kg.

   (a) What is the probability that a jellyfish weighs more than 1.75kg?

   ```
   > 1 - pnorm(1.75, 0.9, 0.5)
   [1] 0.04456546
   ```

   (b) What is the largest weight of jellyfish in the bottom 15% of the population by weight?

   ```
   > qnorm(0.15, 0.9, 0.5)
   [1] 0.3817833
   ```

   (c) Use R to simulate taking a random sample of 10 jellyfish. (Hint:`rnorm`) What is the mean and standard deviation of the jellyfish in your sample?

   ```
   > data<-rnorm(10, 0.9, 0.5)
   > data
    [1]  0.3386087  1.5559924  0.4315115  0.2581337  0.2973410  1.7161218
    [7]  0.5644023  0.7377069  0.6686331 -0.1526290
   > mean(data)
   [1] 0.6415822
   > sd(data)
   [1] 0.5810851
   ```

   (d) Use R to simulate taking a random sample of 100 jellyfish. What is the mean and standard deviation of the jellyfish in your sample?

   ```
   > data<-rnorm(100, 0.9, 0.5)
   > mean(data)
   [1] 0.8827388
   > sd(data)
   [1] 0.4415979
   ```

   (e) Use R to simulate taking a random sample of 1000 jellyfish. What is the mean and standard deviation of the jellyfish in your sample?

```
> data<-rnorm(1000, 0.9, 0.5)
> mean(data)
[1] 0.8930775
> sd(data)
[1] 0.4938173
```

(f) Comment on the values you got from the previous three questions - is this what you expected?

Yes - mean and standard deviation reasonably close to actual values, and get closer as sample size increases.

(g) Use R to take 10 random samples of size 10. What is the average sample mean, and what is the standard deviation of the sample means?

```
> data = numeric(0)
> for (i in 1:10) { data[i] = mean(rnorm(10, 0.9, 0.5)) }
> mean(data)
[1] 0.9033864
> sd(data)
[1] 0.213825
```

(h) Use R to take 10 random samples of size 100. What is the average sample mean, and what is the standard deviation of the sample means?

```
> data = numeric(0)
> for (i in 1:10) { data[i] = mean(rnorm(100, 0.9, 0.5)) }
> mean(data)
[1] 0.8846266
> sd(data)
[1] 0.05046255
```

(i) Use R to take 10 random samples of size 1000. What is the average sample mean, and what is the standard deviation of the sample means?

```
> data = numeric(0)
> for (i in 1:10) { data[i] = mean(rnorm(1000, 0.9, 0.5)) }
> mean(data)
[1] 0.8947339
> sd(data)
[1] 0.01875147
```

(j) Comment on the values you got from the previous three questions - is this what you expected?

Yes - means get closer to the actual mean and standard deviations get smaller.