NAME: SOLUTIONS

**Math 124 Spring 2005: Section 7 TTh 9:35-10:50 AM**
**Final Exam** **Date: May 24, 2005**

**Instructions**: Answer questions 1-8 for full credit. There is a bonus question, but you can get a perfect score without attempting it. Show as much work as you feel reasonable. You have 2 hours 30 minutes. To allow others to fully concentrate at the end please do not leave in the last 5 minutes. You should submit your pages of notes with your test paper.

## Question 1. (*25 points*)

Define each of the following and explain why or how it is used (in the context of an experiment)

(a) *Randomization*

5pts

A procedure for assigning subjects/units/individuals to treatments in a non systematic way. We could give each subject in the population a numbers and then use a computer to generate random numbers to assign the subjects. It is done to reduce bias.

(b) *Confounding*

5pts

Two variables are confounded when their effects on a response variable can't be distinguished from each other. These variables may be either explanatory or lurking. The design of the experiment should seek to avoid unintentionally confounding variables

(c) *Block* and *Block Design*

5pts

A block is a group of experimental units or subjects known to be similar in someway that is expected to affect the response variable. Block designs have ~~treatments~~ randomly assigned within blocks. Blocking is done to reduce variation.

(d) *Single-blind* and *Double-blind*

5pts

Single blind – the recepient (ie patient) does not Know whether (or not) they receive a drug (or placebo).
Double blind – neither the recepient nor the treating physician Knows whether the treatment is the drug or a placebo.
Both are for avoiding Bias

(e) *Replication*

5pts

Repeating the experiment at the same treatment settings many times Done to reduce chance variation

## Question 2. (*25 points*)

An factory manager is interested in the bond strength of a new adhesive product which is being considered for routine use on a production line. The adhesive is used to join together two plastic parts, of which there are unlimited supplies. The adhesive supplier has provided 27 equal sized samples of the adhesive. In the factory the machine which performs the gluing has three different temperature settings (280F, 300F and 320F) and three different pressure settings(100, 150 and 200 psi). A machine which measures the breaking strength is available for use.

(a) Identify the factors, their levels, the treatments and a response variable for this experiment.

Factors – Temperature Settings, Pressure

Levels 280F 300F 320F 100psi 150psi 200psi

treatments: 280F – 100psi
280F – 150psi
280F – 200psi
300F – 100psi
300F – 150psi
300F – 200psi
320F – 100psi
320F – 150psi
320F – 200psi

Response – breaking strength.

12pts

(b) Describe and outline an appropriate design for this experiment.

Randomly assign the 27 glue samples into 9 groups

13pts

| | | Pressure | |
|---|---|---|---|
| | 100psi | 150psi | 200psi |
| 280F | 3glues | 3glues | 3glues |
| Temp 300F | 3glues | 3glues | 3glues |
| 320F | 3glues | 3glues | 3glues |

For further randomization, randomly choose order in which the treatments are applied. After glue has set measure force needed to break bond.

**Question 3.** (*25 points*)

Imagine that you carry out a random experiment where you first roll a fair 6 sided dice, then you roll a fair 4 sided dice.

(a) Give the sample space for this experiment. Then explain what the probability of each individual outcome will be and why.

8pts

S= {(1,1), (1,2), (1,3)(1,4),
(2,1), (2,2), (2,3), (2,4),
(3,1), (3,2), (3,3), (3,4),
(4,1), (4,2), (4,3), (4,4),
(5,1), (5,2), (5,3), (5,4),
(6,1), (6,2), (6,3), (6,4)}

since every outcome is equally likely probability of each individual outcome is $\frac{1}{24}$

(b) Suppose that $X$ is "the square of the larger of the two numbers". Give the probability distribution of this random variable.

| $x$ | 1 | 4 | 9 | 16 | 25 | 36 |
|---|---|---|---|---|---|---|
| $P(X=x)$ | $\frac{1}{24}$ | $\frac{3}{24}$ | $\frac{5}{24}$ | $\frac{7}{24}$ | $\frac{4}{24}$ | $\frac{4}{24}$ |

8pts

(c) What are the mean and standard deviations of $X$?

$$\mu_X = \sum x_i p_i = 1\left(\frac{1}{24}\right) + 4\left(\frac{3}{24}\right) + 9\left(\frac{5}{24}\right) + 16\left(\frac{7}{24}\right) + 25\left(\frac{4}{24}\right) + 36\left(\frac{4}{24}\right)$$

$$= \frac{414}{24} = 17.25$$

9pts

$$\sigma_X^2 = \sum x_i^2 p_i - \mu_X^2 = 1^2\left(\frac{1}{24}\right) + 4^2\left(\frac{3}{24}\right) + 9^2\left(\frac{5}{24}\right) + 16^2\left(\frac{7}{24}\right) + 25^2\left(\frac{4}{24}\right) + 36^2\left(\frac{4}{24}\right) - (17.25)^2$$

$$= \frac{4930}{24} - 17.25^2 = 116.1875$$

$$\sigma_X = \sqrt{116.1875} = 10.7790 \text{ (4dp)}$$

**Question 4.** (*25 points*)

Protein is an important component of both human and animal diets. Although it is well known that grains and legumes contain large amounts of protein, it is not widely recognized that certain grasses can also provide a good source of protein. It is thought that Bermuda grass should contain 20% protein by weight. So one kilogram of Bermuda grass would contain 200g of protein. A scientist wants to verify this claim and so she gathers 75 one-kilogram samples of Bermuda grass and analyzes them for protein content. She finds the mean protein content for her sample is 180g. It is well known that the standard deviation of protein contents from one kilogram samples of grass is 80grams.

(a) Give a 98% confidence interval for the mean protein content of Bermuda grass.

$$180 \pm 2.326\left(\frac{80}{\sqrt{75}}\right)$$

9pts

$$\Rightarrow 180 \pm 2.326(9.238)$$

$$\Rightarrow 180 \pm 21.487$$

$$\Rightarrow (158.513, 201.487)$$

ie between 158.5grams and 201.5grams.

(b) Based upon your confidence interval would it be safe to say that Bermuda grass contains 20% protein by weight? Explain why.

3pts

Yes it would be safe to say that Bermuda grass contains 20% protein because 200g is included in our confidence interval for part (a)

(c) What assumptions are you making when you compute your confidence interval? Which is most important?

5pts

We are assuming that the 1kg grass samples are a SRS from the population of all bermuda grass samples. In addition, we are assuming that protein content is normally distributed in 1kg grass samples. The second assumption is less important because with n=75 samples the CLT tells us that $\bar{x}$ will be approximately normally distributed

## Question 5. (*25 points*)

A computer chip manufacturer claims that at most 2% of the chips it produces are defective. An electronics company purchases a large quantity of chips and wants to determine if this claim is plausible. A sample of 400 of these chips is randomly selected and 13 of the 400 are found to be defective when tested. Suppose that $p$ is the proportion of all chips that are defective.

(a) What is the sample estimate of the proportion of defective chips? What is the standard error of this sample proportion?

5pts

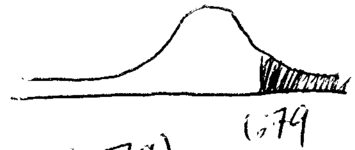
$\hat{p} = \frac{13}{400} = .0325 \quad SE(\hat{p}) = \sqrt{\frac{.0325(1-.0325)}{400}}$

$= .0089 \text{ (4dp)}$

(b) Carry out a hypothesis test of $H_0 : p \leq .02$ against $H_A : p > .02$. Be sure to state your test statistic and its p-value.

9pts

$$Z = \frac{.0325 - .02}{\sqrt{\frac{.02(1-.02)}{400}}} = 1.79 \text{ (2dp)}$$




Since $H_A$ is $p > .02$ $\quad$ Pvalue $= P(Z > 1.79)$

$$= 1 - P(Z < 1.79)$$
$$= 1 - .9633 = .0367$$

(c) Can the computer chip manufacturers claim be supported by this data?

3pts

Since the p-value in the previous part was less that .05 we have some evidence against the null hypothesis. Therefore we would reject the idea that $p \leq .02$ and conclude that the proportion of defectives is above 2%

## Question 6. (*25 points*)

A quality inspector at a widget factory measures the diameter of a specific part of the widget. The factory line must be shut down and re-adjusted if the diameter of this part becomes too small. Each hour the inspector takes a simple random sample of 100 widgets and if the mean diameter is below 5mm then the parts being produced are of poor quality. Suppose that for a particular hour the inspector calculates a sample mean of 4.96mm and sample standard deviation 0.2mm for the 100 widgets sampled.

(a) What is SE $(\bar{x})$?

$$SE(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{.2}{\sqrt{100}} = .02 \checkmark$$

6pts

(b) State appropriate $H_0$ and $H_A$ (null and alternative hypotheses) to test whether or not acceptable parts are being produced.

Let $\mu$ = mean diameter of widget this hour

6pts

$H_0: \mu \geq 5$ (ie machine is fine)

$H_A: \mu < 5$ (ie parts have too small a diameter)

(c) Compute the test statistic and its P-value.

$$t = \frac{4.96 - 5}{.2/\sqrt{100}} = \frac{-.04}{.02} = -2 \qquad df = 100 - 1 = 99$$

9pts

Since $H_A$ is $\mu < 5$ P value $= P(T < -2)$

$= P(T > 2)$ (by symmetry)

using df = 80 line

$1.990 < 2 < 2.088 \Rightarrow .02 <$ P value $< .025$

$\Rightarrow .025 > P(T > 2) > .02$

(d) Would you recommend that the production line be shut down? Explain your answer.

4pts

Since the pvalue in part (c) is small we have sufficient evidence to reject $H_0$ in favor of $H_A$ - ie we have evidence that the mean diameter is below 5mm so the factory line should be shut down and readjusted.

**Question 7.** (*25 points*)

Lead is a pollutant that can have harmful effects on humans. One method of measuring exposure to lead is to examine the lead content of human hair. One of the most common sources of lead exposure is lead-based paint which was widely used up to the 1940's. In 1978, paint containing harmful levels of lead were banned from use on residences, furniture and toys. A research is interested in determining whether or not lead levels have changed from in the past. They gather a dataset which contains hair lead measurements in micrograms for adults who died between 1880 and 1920 and modern adults. The following table summarizes the results

| Group | Size | Mean | Standard Deviation |
|---|---|---|---|
| 1880-1920 | 30 | 48.5 | 14.5 |
| Modern | 100 | 26.6 | 12.3 |

You may assume that $\mu_1$ is the mean lead level for adults in the 1880-1920 period and $\mu_2$ is the mean lead level for modern day adults.

(a) What is the standard error of the difference between the two sample means $\bar{x}_1$ and $\bar{x}_2$?

9pts

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{14.5^2}{30} + \frac{12.3^2}{100}} = \sqrt{8.5212}\ (4dp)$$
$$= 2.9191\ (4dp)$$

(b) Do modern adults have lower lead levels? State appropriate $H_0$ and $H_A$ (null and alternative hypotheses) for answering this question.

8pts

$H_0: \mu_1 - \mu_2 \le 0$ (ie modern adults have higher or same levels)

$H_A: \mu_1 - \mu_2 > 0$ (ie modern adults have lower levels)

(c) Carry out this test. Be sure to state your test statistic value, its degrees of freedom and report your P-value. State in words what you conclude based upon your hypothesis test.

10pts

$$t = \frac{48.5 - 26.6}{\sqrt{8.5212}} = 7.502\ (3dp)$$

7.502

$df = \min(29, 99) = 29$

from t-table

Since $H_A$ is $\mu_1 - \mu_2 > 0$ Pvalue $= P(T > 7.502)$

$3.657 < 7.502 \Rightarrow .0005 > P(T > 7.502)$

Since Pvalue is so small reject $H_0$. Conclude that modern adults have lower levels.

## Question 8. (*25 points*)

Two insect sprays are to be compared for effectiveness at killing insects. Two rooms of equal size are sprayed. One with spray 1 and the other with spray 2. Then 100 insects are then released into each room. After 2 hours the number of dead insects are counted. In the room sprayed with spray 1 a total of 64 dead insects were found. In the other room 52 dead insects were found.

(a) Find a 96% confidence interval for the difference in proportion of insects killed be each spray.

Let $p_1$ = proportion of insects sprayed with spray 1 killed
$p_2$ = proportion of insects sprayed with spray 2 killed

$\hat{p}_1 = \frac{64}{100} = .64$ $\hat{p}_2 = \frac{52}{100} = .5250$ a 96% CI for $p_1 - p_2$ is

15 pts

$.64 - .52 \pm 2.054\sqrt{\frac{.64(1-.64)}{100} + \frac{.52(1-.52)}{100}}$

$\Rightarrow 0.12 \pm 0.1423$ (4dp)

$\Rightarrow (-.0223, .2623)$

(b) Based upon your confidence interval or otherwise determine whether or not the two sprays differ in effectiveness.

Since 0 is inside the CI computed in the previous step a 4% level of significance test would not reject the null hypothesis

10 pts

$p_1 - p_2 = 0$ (ie $p_1 = p_2$) therefore based on this data we can not conclude that the sprays differ in effectiveness