Statistics 13 – Midterm Review Guide (Fall 2005)

http://www.stat.ucla.edu/%7Edinov/courses_students.html

I. Problem set I:

Suppose that in a population of American adult coffee drinkers, the change in measured heart rate ten minutes after drinking 6 ounces of coffee is normally distributed with a mean increase of 7.3 beats per minute and a standard deviation of 11.1 beats per minute. A negative change would indicate a decrease in heart rate. In answering the following questions, ignore the reality that heart rate measurements are discrete (number of beats in a fixed time interval) and imagine that we used instead a more sophisticated form of measurement that gives continuous values.

(a) What proportion of individuals in the population would have a measured decrease in heart rate?

Solution: The answer is the proportion of individuals with measurements below 0. Measurements are normal with $\mu = 7.3$ and $\sigma = 11.1$. Let Y be the measurement of an individual.
$P(Y < 0) = P(Z < -0.66)$
$= 0.2546$

(b) The middle 90 percent of individuals have measured changes between what two values?

Solution: From the table, the 0.95 quantile is between 1.64 and 1.65. I will use 1.645. Ninety percent of individuals have measurements between $7.3 - 1.645(11.1) = -11$ and $7.3 + 1.645(11.1) = 25.6$.

(c) In a random sample of eight adult American coffee drinkers, what is the probability that the sample mean change in heart rate would be below zero?

Solution: The sampling distribution of $\bar{Y}$ is normal with mean 7.3 and standard deviation 11.1/p8 = 3.924. The area to the left of zero is the following.
$P(\bar{Y} < 0) = P(Z < -1.86)$
$= 0.0314$

(d) In a random sample of eight individuals from the population, what is the probability that exactly seven of eight individuals in the sample have increases in their measured heart rates?

Solution: If X is the number of individuals in the sample whose heart rate increases, then X has a binomial distribution with
$n = 8$ and $p = 1 - 0.2546 = 0.7454$.
$P(X = 7) = 0.2604$

(e) Use the normal approximation to the binomial distribution to estimate the probability that there would be 160 or more individuals whose heart rate increased in a random sample of 200 individuals from the population.

Solution: If the sample size is 200, then X is binomial with $n = 200$ and $p = 0.7454$. The mean is $\mu = 149.08$ and the standard deviation is 6.1608. The exact answer is

P (X _ 160) = 0.0428

but you need a computer to get this answer. The normal approximation (with the correction for continuity) is the area to the right of 159.5. If Y is normal with $\mu = 149.08$ and $\sigma = 6.1608$,

$P(Y > 159.5) = P(Z > 1.69)$

$= 0.0455$

Without the continuity correction, this would be as follows.

$P(Y > 160) = P(Z > 1.77)$

$= 0.0384$

II. Problem set II

The purple pigment anthocyanin causes flower petals to be purple. The biochemical pathway for the synthesis of anthocyanin from a colorless precursor requires three enzymes — if any of the three enzymes is not produced, then the biochemical pathway is broken, no purple pigment is produced, and flower color is white. Each of the three enzymes is responsible for one step of the process and is controlled by a single gene with multiple alleles, some of which produce an effective enzyme and some of which do not. We simplify the problem by considering genetic crosses among individuals that all produce one of the enzymes, but have multiple alleles for the other two genes. Genes C and P each have two alleles (C and c, P and p). Individuals with genotypes CC or Cc produce enzyme C, but individuals with genotype cc do not. For gene P, the situation is identical — only the homozygous recessive genotype pp fails to produce enzyme P. You may assume that the two genes are inherited independently. An individual plant with both enzymes has purple flowers, while an individual lacking one enzyme or both has white flowers.

These questions are all based on the following set of genetic crosses. In the first cross, Individual A with genotype CCPp is crossed with Individual B with genotype ccPp. Standard genetic theory shows that the offspring can have three possible genotypes, CcPP, CcPp, and Ccpp with probabilities 1/4, 1/2, and 1/4, respectively. In a second cross, a randomly selected offspring from the first cross, Individual C, is crossed with Individual D with genotype Ccpp producing Individual E. In solving each of the following problems, present sufficient work so that your method of solution is clear. Clearly define any notation for events you introduce for the problem. (For example, you might specify the event

Epurple = {Individual E has purple flowers}.)

Answer the following questions and show your work.

(a) If Individual C has genotype CcPP, what is the probability that Individual E can produce enzyme C?

(b) If Individual C has genotype CcPp, what is the probability that Individual E has white flowers?

(c) What is the probability that Individual E has white flowers?

(d) What is the probability that Individual E has white flowers given that Individual C has purple flowers?

(e) What is the probability that Individual C has purple flowers given that Individual E has white flowers?

Solution: A tree diagram is useful to guide these calculations.
CcPP
CcPp
Ccpp
0.25
0.5
0.25
P
W
P
W
P
W
0.75
0.25
0.375
0.625
0
1
0.1875
0.0625
0.1875
0.3125
0
0.25
Individual C Individual E

(a) E can produce enzyme C if its genotype for gene C is either CC or Cc. Each of its parents has genotype Cc. We have that

P (E produces C | C is Cc) = 1 − P (E does not produce C | C is Cc)
= 1 − P (C gives c and D gives c | C is Cc)

= 1 − (1/2)(1/2)
= ¾

(b) E has white flowers if it cannot produce enzyme C or it cannot produce enzyme P. Let EW be the event that E has white flowers and EP be the event that E has purple flowers.

P (EW | C is CcP p) = 1 − P (EP | C is CcP p)
= 1 − P (E has enzyme C and E has enzyme P | C is CcP p)
= 1 − P (E has enzyme C | C is CcP p) × P (E has enzyme P | C is CcP p)
= 1 − (3/4)(1/2)
= 5/8

(c) The probability that E has white flowers is a weighted average of the conditional probabilities of white flowers weighted by the probabilities of the genotype of C.

P (EW) = (1/4)(1/4) + (1/2)(5/8) + (1/4)(1) = 5/8

(d) We find the probability that E has white flowers given that C has purple flowers by using the rule for conditional probabilities. Let CP be the event that Individual C has purple flowers.

P (EW |CP) = P (EW and CP)
P (CP)
= (1/4)(1/4) + (1/2)(5/8) 3/4

(e) Nominally, this is Bayes' Theorem, but we can use previous work to answer it.
P (CP |EW) = P (CP and EW)
P (EW)
=
(1/4)(1/4) + (1/2)(5/8) 5/8
= 3/5

III. Problem set III

A zoologist measured the heights of 12 randomly selected giraffes. The sample mean *height* was 16.5 ft with standard deviation 3.8 ft.

Problem. Based on this data, what is the standard error of the sample mean?

a) 1.20 b) 1.10 c) 0.317 d) 4.17 e) 4.07

Problem. Assuming the above sample standard deviation is close to the true population standard deviation, what would you expect the standard error of the mean would be for the heights of 150 new randomly selected giraffes?

a) 0.0963 b) 0.0271 c) 0.0253 d) 1.18 e) 0.310

Problem. It is known that the population standard deviation of seal weights is 10 pounds. A study is being planned in which the desired standard error of the sample mean should be no more than 3 pounds. How many seals should be included in the study? (Be careful how you round.)

a) 9 b) 13 c) 11 d) 12 e) 10

Problem. A dentist asks 20 randomly selected patients if they use *Listerine* mouth wash, and finds that 7 do use it. Based on this data, construct a 92 percent confidence interval for the true proportion p of patients who use *Listerene*.

a) (-0.578, 1.32)
b) (0.050, 0.516)
c) (-0.454, 1.19)
d) (-0.347, 1.09)
e) (0.194, 0.546)

Problem. In a study two independent random samples of subjects were assigned to Diet A and Diet B. After a year, their cholesterol levels were recorded. The sample average cholesterol level for the 10 subjects on Diet A was x_A = 4.32 mmol/L with sample standard deviation s_A = 0.81 mmol/L. The sample average cholesterol level for the 9 subjects on Diet B was x_B = 4.91 mmol/L with sample standard deviation s_B = 0.80 mmol/L. Assume that the cholesterol levels are normally distributed. Let μ_A denote the **true** mean cholesterol level for people on Diet A and μ_B denote the **true** mean cholesterol level for people on Diet B. What is the value of the test statistic for a hypothesis test of H_0: $\mu_A = \mu_B$ against the alternative H_1: $\mu_A != \mu_B$?

a) -1.60 b) -1.83 c) -1.25 d) -1.86 e) -1.71

Data from two independent random samples gave the following results:

	Sample1	Sample2
n	9	15

sample_avg	5.3	5.7
SE_Sample_Avg	1.2	2.1

Problem. Based on this data, what is the sample standard deviation for sample 1?

a) 2.42 b) 10.8 c) 0.133 d) 3.60 e) 1.11

Problem. Based on this data, what is the standard error of the difference of the two sample averages ($\overline{Y}_1 - \overline{Y}_2$)?

a) 2.99 b) 1.67 c) 4.31 d) 1.82 e) 2.42

Problem. A researcher is planning to compare the average ear lengths of two species of rabbits. She intends to use a two-tailed t-test at the 10 percent significance level. She can afford to collect 52 observations in each of the two groups in her study. What is the smallest effect size for which she has at least 95 percent power? Show your work! Verify your answer using the SOCR power Applet: http://socr.stat.ucla.edu/Applets.dir/PowerApplet.html.

a) 0.30 b) 3.0 c) 0.90 d) 0.65 e) 2.4

Problem. Researchers wish to test the null hypothesis that the population mean weight of wild pandas is the same as that for pandas in captivity against the alternative that the means are not equal. It is known that the weights of pandas in the wild and in captivity are normally distributed. A two sample t-test is conducted, and the resulting test statistic is 1.109 with 11 degrees of freedom. Which of the following are the most accurate bounds on the p-value?

a) 0.05 < p-value < 0.06
b) 0.025 < p-value < 0.03
c) 0.10 < p-value < 0.20
d) 0.20 < p-value < 0.40
e) p-value < 0.0005

Problem. What is the 98^{th} percentile of the t-distribution with 27 degrees of freedom?

a) 2.845 b) -2.158 c) -2.539 d) 2.861 e) -2.845 f) 2.539 g) -2.861 h) 2.158

Problem. The top running speeds of 2 randomly selected Chipmunks have a sample mean of 1.3 mph with sample standard deviation 0.2 mph. Assume the top running speeds for Chipmunks have a normal distribution. Based on the data, construct a 90 percent confidence interval for the true mean top running speed of Chipmunks.

a) (0.887, 1.71)
b) (1.14, 2.00)
c) (0.407, 2.19)
d) (1.07, 1.53)
e) (0.693, 1.91)

For each of the following situations, suppose $H_0 : \mu_1 = \mu_2$ is being tested against $H_A : \mu_1 \mathrel{!}= \mu_2$. State whether or not H_0 would be rejected in each case.

Problem. p-value=0.71, $\alpha = 0.10$.

a) do not reject H_0 b) reject H_0

Problem. $t_s = 1.121$ with 12 degrees of freedom, $\alpha = 0.10$.

a) reject H_0 b) do not reject H_0

Problem. 90 percent confidence interval (0.045, 1.12), $\alpha = 0.10$.

a) do not reject H_0 b) reject H_0

Problem. Researchers wish to test the null hypothesis that the population mean carboxyhemoglobin level is the same for smokers as it is for nonsmokers against the alternative that smokers have a LARGER population mean carboxyhemoglobin level. It is known that carboxyhemoglobin levels are normally distributed for both smokers and nonsmokers. A two sample t-test is conducted, and the resulting test statistic is 2.109 with 17 degrees of freedom. Which of the following are the most accurate bounds on the p-value?

a) $0.04 < \text{p-value} < 0.05$
b) $0.05 < \text{p-value} < 0.06$
c) $0.025 < \text{p-value} < 0.03$
d) $0.10 < \text{p-value} < 0.20$
e) $0.05 < \text{p-value} < 0.10$

Problem. Suppose a similar study is conducted and a random sample of 54 smokers has a sample mean carboxyhemoglobin level of 6.1 with sample standard deviation 1.4. A random sample of 61 nonsmokers, independent of the random sample of smokers, has a sample mean carboxyhemoglobin level of 5.3 with sample standard deviation 1.1. What is the value of the test statistic for the one-sided hypothesis test?

a) 3.02 b) 3.80 c) 4.36 d) 4.85 e) 3.38

Problem. Which of the following is a valid 90 percent confidence interval for the true mean difference in carboxyhemoglobin level? (Use 100 degrees of freedom.)

a) (0.651, 0.949)
b) (0.692, 0.908)
c) (-0.195, 1.79)
d) (0.407, 1.19)
e) (0.794, 0.806)

Problem. 99, 95, 90, 80, and 60 percent confidence intervals were constructed for the difference in two population means using the t procedures on the same data set. The following intervals are those confidence intervals. Which is the 95 percent confidence interval?

a) (-1.97; -0.628)
b) (-4.24; 1.64)
c) (-3.18; 0.577)
d) (-2.38; -0.223)
e) (-2.77; 0.171)

Problem. The sample standard deviation of a set of n measurements is 8, and the standard error of the sample mean is 2. What is the size of the sample?

a) 14 b) 21 c) 15 d) 16 e) 13

IV. Problem set IV

Problem. In a simple random sample of 20 gymnasts it is found that 12 of them have had knee surgery. Find a 90 percent confidence interval for the proportion of all gymnasts who have had knee surgery.

a) (-0.153, 1.33)
b) (-0.385, 1.56)
c) (-0.0676, 1.24)

d) (0.418, 0.758)
e) (-0.187, 1.36)

Problem. A researcher is planning to compare the average jumping distance of two species of kangaroos. He plans to analyze the data with one-tailed t-test at the 5 percent significance level. The researcher plans to use 6 kangaroos in each group. Supposing that the true difference in population means is 2 feet and the standard deviation of each population is 0.8 feet, what is the probability that the difference will be detected?

a) 0.50 b) 0.90 c) 0.80 d) 0.99 e) 0.95

Problem. A researcher is planning to compare the average sleeping time of two species of bats. She will be analyzing the data with a one-tailed t-test. She expects the difference in the true population mean sleep times is 1 hour, and that the standard deviation of each population is 0:5 hours. For 80 percent power at a 1 percent significance level, how many bats should be included in each sample?

a) 7 b) 5 c) 17 d) 12 e) 11

Problem: Compute the expected value and the standard deviation for the random variable X = total sum of the outcome or rolling 2 fair six-face dice. Would you bet on X < 6 if winning returns your original bet plus twice the same amount, and when loosing (X>=6) the house collects your bet? Explain!

Problem. Lung volume can be measured by exhaling into a simple measuring device. These measurements are not precise and repeated measurements will yield different values. These repeated measurements are known to have a normal distribution whose mean is the true lung volume. Dave measures his lung volume on 5 different days and gets the values: 101, 111, 119, 105, 107 (in cubic inches). Construct a 99 percent confidence interval for George's true lung volume.

a) (91.2, 108.6)
b) (93.4, 123.8)
c) (94.5, 122.7)
d) (91.9, 107.9)
e) (93.9, 123.3)

Problem. A researcher is planning to compare the average jumping distance of two species of frogs. The data will be analyzed with a one-tailed t-test at the 1 percent significance level. He can afford to collect 20 observations in each of the two groups in his study. It is known that the true population standard deviation for the jumping distances of both species of frog is 0.7 feet. What is the smallest difference in population average jumping distances that can be detected with 90 percent power? [**Hint**: $N = 2\sigma^2 (z_{\alpha/2} + z_{\beta})^2 / \Delta^2$, where N is the sample-size in each group, σ = the common standard deviation of each group, Δ = the smallest difference you want to detect with high power, α = significance level, and β = *1-Power*.] Validate using http://socr.stat.ucla.edu/Applets.dir/PowerApplet.html.

a) 0.08 b) 0.41 c) 0.72 d) 0.84 e) 0.98

A random sample, Sample 1, is taken from Population 1. Another random sample, Sample 2, is taken from Population 2, and it is independent of Sample 1. The data set shown below:

Sample	1	1	9	12	15	15	16	17	40	65
Sample	2	3	7	8	11	15	18	19	44	

Problem. What is the value of the T test statistic? Interpret this value!

Problem. Which statement below is a valid alternative hypothesis for which the data differs in the direction of the alternative?

a) Sample 1 tends to have larger values than Sample 2.
b) Population 1 tends to have smaller values than Population 2.
c) Population 1 tends to have larger values than Population 2.
d) Sample 1 tends to have smaller values than Sample 2.

Problem: Suppose 9% of children are born before 36 weeks of gestation (that is one month or more premature). A child born premature has a 70% of being low birth weight (weighing less than 2500 grams at birth). A child born at full term has a 15% chance of being low birth weight. What is the probability that a newborn is of low birth weight? What is the probability that a baby weighing 2000 grams was born premature?