22S:30/105, Statistical Methods and Computing
Spring 2012, Instructor: Cowles
Midterm 2

Name: Solutions Course no. (30 or 105) _____

Show your work on any problems that involve calculations.

I will grade on a curve and will give partial credit wherever possible.

1. According to your textbook (p. 264), the probability model for the blood type of a randomly chosen person in the U.S is:

Blood type	O	A	B	AB
Probability	0.45	0.40	0.11	?

(a) This probability model is (circle one):

i. ~~continuous~~
ii. discrete
iii. equally likely
iv. none of the above
v. all of the above

2 (b) Assuming that there are no blood types other than those listed in the table, what is the probability that a randomly selected American has type AB blood?

$1 - (0.45 + 0.40 + 0.11) = .04$

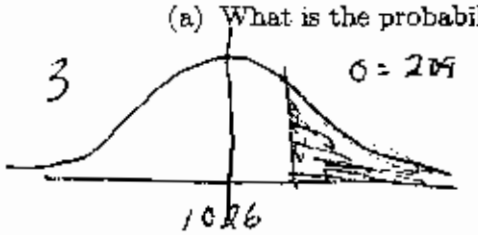
2 (c) What is the probability that a randomly selected American does not have Type O blood?

$1 - 0.45 = 0.55$

2. The total SAT scores of high school seniors in recent years have a roughly normal distribution with mean $\mu = 1026$ and standard deviation $\sigma = 209$.

(a) What is the probability that a randomly selected SAT score is higher than 1100?

3

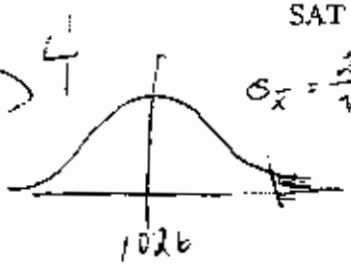


$P(X > 1100) =$

$P\left(Z > \frac{1100 - 1026}{209}\right) = P(Z > .354)$

$= 1 - P(Z < .354) = 1 - 0.638 = 0.362$

(b) What is the probability that the sample mean $\bar{x}$ from a random sample of 70 SAT scores is higher than 1100?



$\sigma_{\bar{x}} = \frac{209}{\sqrt{70}} = 24.98$

$P(\bar{X} > 1100) =$

$P\left(Z > \frac{1100 - 1026}{24.98}\right) = P(Z > 2.962)$

$= 1 - P(Z < 2.962) = 1 - .9985 = .0015$

12

3. The primary purpose of a research study is to estimate the mean time that it takes patients who undergo coronary artery bypass graft (CABG) surgery to regain consciousness after surgery. Both a point estimate and a 99% confidence interval are desired. The investigators plan to draw a simple random sample of patients who had CABG from a huge database. Time to regaining consciousness is not recorded in the database, so it will have to be abstracted from the patients' hospital records.

(a) What is the population of interest in this study? (Circle one)

i. all patients who ever undergo CABG

ii. the patients who are selected at random from the database

iii. the mean time to regaining consciousness among all CABG patients

iv. the mean time to regaining consciousness among the patients who who selected at random frin tge database

(b) What is the parameter of interest in this study? (Circle one)

i. all patients who ever undergo CABG

ii. the patients who are selected at random from the database

iii. the mean time to regaining consciousness among all CABG patients

iv. the mean time to regaining consciousness among the patients who who selected at random frin tge database

(c) Is it likely that the distribution of time to regaining consciousness follows a normal distribution among all CABG patient? Briefly state why or why not.

No. The distribution is likely to be right skewed. Time to regain consciousness cannot be less than 0. A few patients could take a very long time.

(d) Suppose that the investigators strongly believe that the distribution of time to regaining consciousness among all CABG patients follows a normal distribution with standard deviation $\sigma = 1.5$ hours.

If the investigators want to obtain a 99% confidence interval of width no greater than 0.5 hours, how many patients should they enroll in their study? (Numeric answer; show your work.)

$m = \frac{1}{2}$ width of interval

$n = \frac{2.58(1.5)^2}{(0.25)} = 239.63$

round up to 240 patients

2 pts if use $m = 0.5$

(e) Suppose that a secondary purpose of the study is to determine whether the mean time to regaining consciousness for all CABG patients is less than 8 hours.

i. Write the null and alternative hypotheses, using the standard symbols from your textbook and lectures.

$H_0: \mu \geq 8$ hrs

$H_A: \mu < 8$ hrs

ii. Which of the following can a test of significance do? (Circle all that apply.)

A. prove that a null hypothesis is false

B. assess the evidence contained in data against a null hypothesis in favor of an alternative hypothesis

C. determine the probability that a null hypothesis is true

D. all of the above

iii. Suppose that the z statistic for the above hypothesis test turned out to be 2.01.

A. Determine the p-value of the test. (Numeric answer; show your work.)

H_A: $\mu < 8$ hrs, a negative z statistic is evidence for H_A; p-val $= Pr(Z \leq -2.01) = .977$

B. Would the results be significant at the 0.01 level? (yes/no) Briefly justify your answer.

p-val $= .977 > \alpha = .01$

significant $\iff$ reject H_0

4. In a study conducted in Italy, 10 patients with hypertriglyceridemia were placed on a low-fat, high-carbohydrate diet. Before the start of the diet, cholesteral and triglyceride measurements were recorded for each subject. The variables in the dataset "cholesterol.dat" are

- patient number
- cholesterol level (mmol/l)
- triglyceride level (mmol/l)

The researchers were interested in using the pre-diet data to estimate the mean cholesterol level in all untreated patients with hypertriglyceridemia. Refer to the SAS code and output below in answering the following questions:

(a) The type of confidence interval computed by `proc means` is (circle one):

i. p interval

ii. t interval

iii. z interval

iv. none of the above

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

(b) The assumptions for the use of this type of confidence interval are (circle all that apply):

i. The sample is a simple random sample from the population of interest.
ii. The population distribution is approximately normal.
iii. The population distribution is approximately t.
iv. The population mean is known.
v. The population standard deviation is known
vi. None of the above

(c) Circle all statements below that represent valid interpretations of the confidence interval.

i. 95% of patients with hypertriglyceridemia have cholesterol levels between 5.615 and 7.851.
ii. We are 95% confident that the sample mean lies between 5.615 and 7.851.
iii. We are 95% confident that the population mean lies between 5.615 and 7.851.
iv. There is 95% probability that the sample mean lies between 5.615 and 7.851.
v. There is 95% probability that the population mean lies between 5.615 and 7.851.
vi. None of the above

```
options linesize = 72 ;

data cholesterol ;
infile '/group/ftp/pub/kcowles/datasets/cholesterol.dat' ;
input patno cholesterol triglycerides ;
run ;

proc means data = cholesterol n mean clm alpha = .05;
var cholesterol  ;
run ;
```

Analysis Variable : cholesterol

N	Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
10	6.7330000	5.6146762	7.8513238