Statistics 1040, Sections 002, 003 & 004, Quiz 1 (20 Points)

January 18, 2002

Your Name: Solution

Question 1: Observational Studies and Experiments (**10 Points**)

The Public Health Service studied the effects of smoking on health, in a large sample of representative households. For men and for women in each age group, those who had never smoked were on average somewhat healthier than the current smokers, but the current smokers were on average much healthier than those who had recently stopped smoking. Answer the following three questions:

(2) 1. Is this an (observational study) or a contolled experiment? Circle your answer.

(4) 2. Why did they study men and women and the different age groups separately?

answer from "Stat 1040" workbook:

"They studied the groups separately to eliminate the effects of the confounding factors of age and gender."

-2 if "confounding factor" not mentioned

-1 (for each) if it is not clearly stated that age + gender are confounding factors

(4) 3. The lesson seems to be that you shouldn't start smoking, but once you have started, don't stop. Comment briefly.

answer from "Stat 1040" workbook:

"That is not an appropriate conclusion because there are confounding factors. For example, those who recently stopped smoking may have done so on doctor's orders because they had severe health problems."

Note: It is not specified what "recently" means in the statement above - a week, a month or a year.

Please turn over!

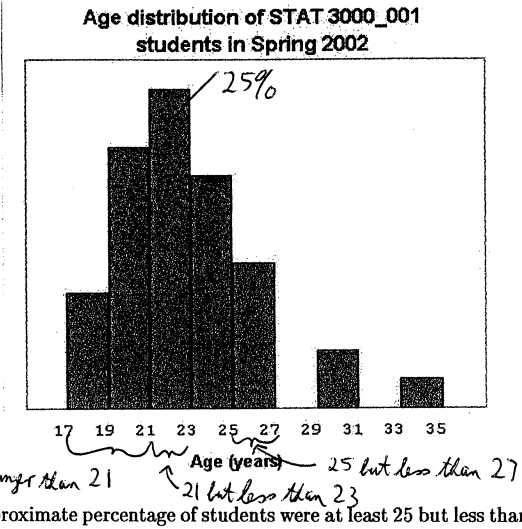
If they stopped "very" recently there might be some additional withdrawal effect since they no longer get nicotine.

If recently means a year or so, this withdrawal effect should have been overcome - but then those people are likely to have a long term heart or lung disease, cancer, etc.

[1]

Question 2: Histograms (**10 Points**)

The histogram below shows the age distribution of Stat 3000, Section 001, students for the Spring 2002 semester. Unfortunately, the labels on the vertical axis have been deleted. However, the instructor recalls that there have been about 25% of students who were at least 21 but less than 23 years old. Try to help the instructor to fill in some of the missing percentages.



-3 if a percentage makes no sense, i.e., is outside the acceptable answers

(5) 1. What approximate percentage of students were at least 25 but less than 27 years old?

answer: about 12%

explanation (not required): the height of the bar from 25 to 27 is slightly less than 1/2 of the height of the bar from 21 to 23, so slightly less than 1/2 · 25%

acceptable answers: anything between 10% and 13%

exact answer: 12.5% (the 21 to 23 group actually contains exactly 27%)

(5) 2. What approximate percentage of students were younger than 21 years of age?

answer: about 30%

explanation (not required): stack the 17 to 19 bar on top of the 19 to 21 bar and this gets higher than the 21 to 23 bar; the extra height represents about 5%

acceptable answers: anything between 26% and 40%

exact answer: 32.5%

Statistics 1040, Sections 002, 003 & 004, Quiz 2 (20 Points)

January 25, 2002

Your Name: _____

Question 1: Measures of Center and Spread (**20 Points**)

Below are the temperatures (in degrees Fahrenheit) for five cities in Alaska on January 4, 2002, as reported in USA Today (see your handouts from the first day of class):

City	Temperature
Anchorage	30
Barrow	-18
Fairbanks	-3
Juneau	36
Nome	0

(6) 1. Find the **average temperature** for these cities in Alaska. **Show your work!**

$$\text{avg} = \frac{30 + (-18) + (-3) + 36 + 0}{5} = \frac{45}{5} = 9$$

-3 if only final result (correct!)
-2 for incorrect sum
-2 for missed value

(4) 2. Find the **median temperature** for these cities in Alaska.

sorted list: -18 -3 0 30 36
↑
center value : 0
median = 0

-3 for incorrect median

Please turn over!

(10) 3. Find the **standard deviation of the temperatures** for these cities in Alaska. **Show your work!**

① avg = 9

② $30 - 9 = 21$
$-18 - 9 = -27$
$-3 - 9 = -12$
$36 - 9 = 27$
$0 - 9 = -9$

③ $21^2 = 441$
$(-27)^2 = 729$
$(-12)^2 = 144$
$27^2 = 729$
$(-9)^2 = 81$

④ $\frac{441 + 729 + 144 + 729 + 81}{5} = \frac{2124}{5} = 424.8$

⑤ $SD = \sqrt{424.8} = 20.61$

-5 if only final result (correct!)
-2 each mistake (not for follow-up mistakes)
-9 if wrong formula only (no values at all)
-2 for missed value
-2 if (avg − entry)
-5 if summation of values (and not squares)
-5 if avg. not subtracted in step ②

Formulas:

$$\text{avg} = \frac{\text{sum of all numbers}}{\text{how many numbers}}$$

$$\text{SD} = \sqrt{\text{average of [(deviations from avg)}^2\text{]}}$$

Statistics 1040, Sections 002, 003 & 004, Quiz 3 (20 Points)

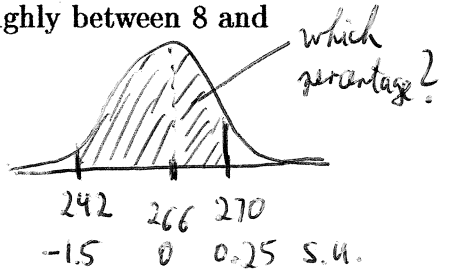
February 1, 2002

Your Name: _____

Question 1: Normal Approximation for Data **(20 Points)**

The length of human pregnancies from conception to birth varies according to a distribution that approximately follows the normal curve. The average is 266 days and the standard deviation is 16 days.

What percentage of pregnancies lasts between 242 and 270 days (roughly between 8 and 9 months)?



Show your work!

① Convert each number into standard units:

$\frac{242-266}{16} = -1.5$ s.u. $\qquad \frac{270-266}{16} = 0.25$ s.u.

②: Find the area (percentage) related to the numbers from ① (use Table):

area from -1.5 to 1.5: 86.64%

area from -0.25 to 0.25: 19.74%

③: Find the area (percentage) of interest:

area from -1.5 to 0: $\frac{1}{2} \cdot 86.64\% = 43.32\%$

area from 0 to 0.25: $\frac{1}{2} \cdot 19.74\% = 9.87\%$

total area from -1.5 to 0.25: $43.32\% + 9.87\% = \underline{53.19\%}$

Answer: about 53.19% of all pregnancies last between 242 and 270 days.

-4 for each incorrect (or missing) s.u. $\qquad$ +4 if correct graph only

-3 for each incorrect table value

-4 for incorrect final result

-2 for each minor (computational) errors

Statistics 1040, Sections 002, 003 & 004, Quiz 4 (20 Points)

February 8, 2002

Your Name: _____

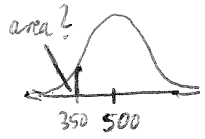
↓ Chapter 5, Review Question 7, page 95

Question 1: Percentiles and the Normal Curve **(12 Points)**

Among freshmen at a certain university, scores on the Math SAT followed the normal curve, with an average of 500 and an SD of 100. Fill in the blanks. **Show your work!**

(6)

1. A student who scored 350 on the Math SAT was at the _____7_____ th percentile of the score distribution.



Transfer 350 in s.u.: $\frac{350-500}{100} = -1.5$

Area between −1.5 to 1.5 (from Table): 86.64%

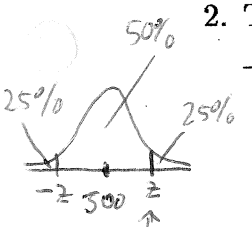
Area below −1.5: $\frac{100\% - 86.64\%}{2} = \frac{13.36\%}{2} = 6.68\%$

i.e., about 7% of the area lies below 350

-2 for incorrect s.u.
-1 for incorrect table value
-2 for incorrect area

(6)

2. To be at the 75th percentile of the distribution, a student needed a score of about _____565 (or 570)_____ points on the Math SAT.



Find a value z such that the area between −z and z is about 50%: z = 0.65 (48.43%) or z = 0.70 (51.61%)

Transfer into original units: for z = 0.65: $0.65 \cdot 100 + 500 = 565$

for z = 0.70: $0.70 \cdot 100 + 500 = 570$

Note: either 0.65 or 0.70 are valid answers.

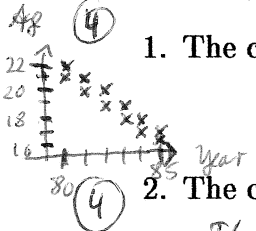
A score of 565 (or 570) was needed to be at the 75th percentile.

-2 for incorrect area
-2 for incorrect z (not 0.65 or 0.70)
-2 for incorrect original units

↓ Chapter 8, Exercise Set B, Question 7, page 130

Question 2: Correlation **(8 Points)**

Investigators are studying registered students at a university. The students fill out questionnaires giving their year of birth, age (in years), age of mother, and so forth. Fill in the blanks, using the options given below, and **explain briefly.**



(4)

1. The correlation between student's age and year of birth is _____(b) nearly −1_____.

Book: "The older you are, the earlier you were born; but there is some fuzz, depending on whether your birthday is before or after the day of the questionnaire."

(4)

2. The correlation between student's age and mother's age is _____(e) somewhat positive_____.

The older the student, the older the mother. However, there is a lot of variation when to give birth (ages between 12 to 50 years are possible).

Options: (a) -1 (b) nearly -1 (c) somewhat negative (d) 0
(e) somewhat positive (f) nearly 1 (g) 1

"slightly" wrong answers: 1
1: (a), (c), (f)
2: (d), (g)

0: wrong answer, no explanation
1: "totally" wrong answer, some explanation
2: "slightly" wrong answer, no explanation
3: "slightly" wrong answer, some explanation
3: correct answer, no explanation

Statistics 1040, Sections 002, 003 & 004, Quiz 5 (20 Points)

March 1, 2002

Your Name: _____

Question 1: The Regression Line **(20 Points)**

For the men of age 18 to 24 years in the HANES sample, the relationship between height and systolic blood pressure can be summarized as follows:

		Average	SD
x	Height	70"	3"
y	Blood pressure	124mm	15mm

The correlation coefficient r is -0.2 and we can assume that the scatterplot is football-shaped.

1. Find the regression equation for predicting blood pressure from height. (10 Points)

slope $= r \cdot \frac{SD_y}{SD_x} = (-0.2) \cdot \frac{15}{3} = (-0.2) \cdot 5 = -1$

intercept $= avg_y - slope \cdot avg_x = 124 - (-1) \cdot 70 = 124 + 70 = 194$

regression equation:

blood pressure $= 194 + (-1) \cdot$ height $= 194 -$ height

or $y = 194 - x$

2. Use the regression equation from part 1. to predict the average blood pressure of men who were 6ft tall (note: 1ft = 12"). (5 Points)

6ft $= 6 \cdot 12" = 72"$

blood pressure $= 194 - 72 = 122$ mm (i.e., the predicted average blood pressure of 6ft tall men)

3. Find the r.m.s. error for predicting blood pressure from height. (5 Points)

r.m.s. error $= \sqrt{1-r^2} \cdot SD_y = \sqrt{1-(-0.2)^2} \cdot 15$

$= \sqrt{1-0.04} \cdot 15 = \sqrt{0.96} \cdot 15 = 14.7$ mm

Please turn over!

Formulas:

$$\text{r.m.s. error} = \sqrt{1-r^2} \times SD_y$$

$$\text{slope} = r \times \frac{SD_y}{SD_x}$$

$$\text{intercept} = avg_y - \text{slope} \times avg_x$$

Grading criteria:

in 1., 2., & 3.: -2 for each calculation error
-2 for each incorrect value used
-2 for swapping x and y

in 1.: -3 for incorrect formula for slope
-3 for incorrect formula for intercept
-2 if no final equation stated
-1 if only part of the equation stated (e.g., 194 - height)

in 2.: -3 for incorrect formula for prediction
-1 if correct result, but according to old method

in 3.: -3 for incorrect formula for r.m.s. error
-2 for using SD_x (instead of SD_y)

Statistics 1040, Sections 002, 003 & 004, Quiz 6 (20 Points)

March 8, 2002

Your Name: _____

Question 1: Chance/Probability **(20 Points)**

1. **(5 Points)** When we roll 2 dice once, what is the probability that their total is 4?

Dice 1	Dice 2	Total
1	3	4
2	2	4
3	1	4

3 combinations have a total of 4; number of combinations: 36

⇝ chance that total of 2 dice is 4: $\frac{3}{36} = \frac{1}{12} = 0.0833 = 8.3\%$

2. **(5 Points)** When we draw 2 cards from a well-shuffled deck of 52 cards without replacement, what is the chance that the first card is a ♡ and the second card is a ◇? 52 cards: 13 ♡, 13 ◇, 26 others

dependent events: chance that 1st is ♡: $\frac{13}{52}$; chance that 2nd is ◇ given that 1st is ♡: $\frac{13}{51}$

⇝ chance that 1st is ♡ and 2nd is ◇: $\frac{13}{52} \cdot \frac{13}{51} = \frac{1}{4} \cdot \frac{13}{51} = \frac{13}{204} = 0.0637 = 6.4\%$ (multiplication rule)

3. **(5 Points)** I roll a die and toss a coin. What is the chance that the die shows "6" and the coin lands on H?

independent events: chance of a "6": $\frac{1}{6}$; chance of a H: $\frac{1}{2}$

⇝ chance of a "6" and a H: $\frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12} = 0.0833 = 8.3\%$ (multiplication rule)

4. **(5 Points)** A coin is tossed six times. Two possible sequences of results are

(i) H T T H T H (ii) H H H H H H

(The coin must land on H or T in the order given; H = heads, T = tails).

Which of the following is correct? Explain:

(a) Sequence (i) is more likely.

(b) Sequence (ii) is more likely.

(c) Both sequences are equally likely.

Chapter 13, Review Question 7, page 235:

Workbook answer:

"(c) is correct. Every possible string of H's and T's is equally likely."

Additional explanation:

1 toss: H, T (chance of 1/2 each)

2 tosses: HH, HT, TH, TT (chance of 1/4 each)

3 tosses: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT (chance of 1/8 each)

6 tosses: $2^6 = 64$ possible sequences, each sequence has identical chance of 1/64 to happen

Grading criteria:

general: -2 for calculation error

1.) -3 for incorrect numerator (e.g., $\frac{2}{36}$ or $\frac{4}{36}$)

-2 for incorrect denominator

2.) -2 if not $\frac{13}{52}$ $(= \frac{1}{4})$ for 1st card

-2 if not $\frac{13}{51}$ for 2nd card

-3 for wrong rule (e.g., "+" instead of ".")

3.) -2 if not 1/6 for die

-2 if not 1/2 for coin

-3 for wrong rule (e.g., "+" instead of ".")

-3 if not conducting last step

4.) -1 for incorrect explanation

-2 for no explanation

-4 for (a) or (b) with explanation

-5 for (a) or (b) without explanation

-1 if correct explanation, but nothing marked

Statistics 1040, Sections 002, 003 & 004, Quiz 7 (20 Points)

March 15, 2002

Your Name: _____

Question 1: Box Models, EV, and SE (**16 Points**)

A quiz has 25 multiple choice questions. Each question has 5 possible answers, one of which is correct. A correct answer is worth 4 points, but a point is taken off for each incorrect answer. A student answers all the questions by guessing at random.

1. **(4 Points)** Find the box model.

Number of draws: 25

-2 for minor error
-3 for major error
-1 if number of draws not stated

2. **(6 Points)** Find the expected value, i.e., the number of points a student would get when answering all questions by guessing.

$\text{box average} = \frac{4 + 4 \cdot (-1)}{5} = \frac{0}{5} = 0$

$EV_{sum} = 25 \cdot 0 = 0$

-1 each calculation error
-2 for each minor error
-3 for each major error

3. **(6 Points)** Find the standard error.

$\text{box SD} = \sqrt{\frac{(4-0)^2 + 4 \cdot (-1-0)^2}{5}}$

$= \sqrt{\frac{16 + 4 \cdot 1}{5}} = \sqrt{\frac{20}{5}} = \sqrt{4} = 2$

$SE_{sum} = \sqrt{25} \cdot 2 = 5 \cdot 2 = 10$

-1 each calculation error
-2 for each minor error
-3 for each major error

Please turn over!

Question 2: Law of Averages (**4 Points**)

A box contains 10,000 tickets: 4,000 [0] 's and 6,000 [1] 's. And 10,000 draws will be made at random with replacement from this box. Which of the following best describes the situation, and why? **Explain briefly.**

1. The number of 1's will be 6,000 exactly.
2. The number of 1's is very likely to equal 6,000, but there is also some small chance that it will not be equal to 6,000.
3. The number of 1's is likely to be different from 6,000, but the difference is likely to be small compared to 10,000.

Chapter 16, Review Question 1, page 285:

Workbook answer:

"Option (3.) is the best because the chance error is not likely to be exactly zero, but it is likely to be small compared to the number of draws."

⇝ Law of Averages

-1 for correct answer, no explanation
-2 for 2. with explanation
-3 for 1. with explanation
-4 for 1. or 2., no explanation

Formulas:

$$\text{box average} = \frac{\text{sum of all numbers in box}}{\text{how many numbers in box}}$$

$$\text{box SD} = \sqrt{\text{average of [(deviations from box average)}^2\text{]}}$$

$$EV_{sum} = \text{number of draws} \times \text{box average}$$

$$SE_{sum} = \sqrt{\text{number of draws}} \times \text{box SD}$$

Statistics 1040, Sections 002, 003 & 004, Quiz 8 (20 Points)

March 22, 2002

Your Name: _____

Question 1: EV, SE, and Normal Curve (**20 Points**)

According to government data, 21% of American children under the age of six live in households with income less than the official poverty level. A study of learning in early childhood chooses a sample of 300 children.

1 = child lives in poverty
0 = child does not live in poverty

1. **(4 Points)** Find the box model.

21 × 1	79 × 0

number of draws = 300

-1 if slightly incorrect number of 0/1's in box
-2 if box given as 10 1 (etc.)
-3 for box with something else than 0/1's
-1 if number of draws missing or incorrect

2. **(8 Points)** Find the expected number of children in the sample who come from poverty–level households? And what is the corresponding SE?

box average = $\frac{21}{100} = 0.21 = 21\%$

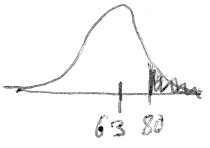
box SD = $\sqrt{\frac{21}{100} \cdot \frac{79}{100}} = \sqrt{0.21 \cdot 0.79} = \sqrt{0.1659} = 0.407$

$EV_{sum} = 300 \cdot 0.21 = 63$

$SE_{sum} = \sqrt{300} \cdot 0.407 = 17.32 \cdot 0.407 = 7.05$

-1 for each calculation error
-1 for each minor mistake
-2 for each major mistake

3. **(8 Points)** Using the normal curve, find the chance that at least 80 of the children in the sample live in poverty.



s.u.: $\frac{80-63}{7.05} = \frac{17}{7.05} = 2.41$

area between −2.40 and 2.40: 98.36%

area above 2.40: $\frac{100\% - 98.36\%}{2} = \frac{1.64\%}{2} = 0.82\%$

There is a chance of about 0.82% that at least 80 children in the sample live in poverty.

-1 for each calculation error
-2 for incorrect curve parameters, i.e., anything else than EV and SE
-2 for incorrect s.u.
-2 for incorrect table value
-2 for incorrect area under the curve

Please turn over!

Statistics 1040, Sections 002, 003 & 004, Quiz 9 (20 Points)

April 5, 2002

Your Name: _____

Question 1: $EV_{\%}$, $SE_{\%}$, and Normal Curve **(20 Points)**

A greenhouse has a large number of petunia seedlings of which 25% are purple, the rest white. If I buy 144 of these seedlings, chosen at random, find the chance that I get at least 30% purple ones. **Show your work!**

1: purple
0: white

box: [25 × 1 75 × 0]

number of draws: 144

each calculation error (-2)
incorrect box (-3)

box avg $= \frac{25}{100} = 0.25$

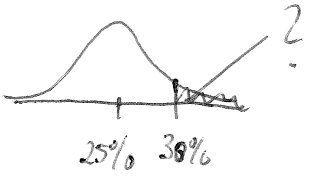
box SD $= \sqrt{\frac{25}{100}\cdot\frac{75}{100}} = \sqrt{0.1875} = 0.433$

$EV_{sam} = 144 \cdot 0.25 = 36$

$SE_{sam} = \sqrt{144} \cdot 0.433 = 5.2$

$EV_{\%} = 25\%$ (5)

$SE_{\%} = \frac{5.2}{144} \cdot 100\% = 3.6\%$ (5)



s.u.: $\frac{30\% - 25\%}{3.6\%} = 1.39$ (4)

area from −1.40 to 1.40: 83.85% (3)

area above 1.40: $\frac{100\% - 83.85\%}{2} = 8.075\% \approx \boxed{8.1\%}$ (3)

Please turn over!

There is a 8.1% chance to get at least 30% of purple petunias

Statistics 1040, Sections 002, 003 & 004, Quiz 10 (20 Points)

April 12, 2002

Your Name: _____

Question 1: Confidence Intervals for Percentages **(20 Points)**

1) (15 Points) We take a random sample of 500 Cache Valley drivers and find that 346 always wear a seat belt. Find a 95% confidence interval for the percentage of *all* Cache Valley drivers who always wear a seat belt. **Show your work!**

From: Stat 1040, Spring 2000, Final Test, Friday, May 5, 2000, Question 3

-1 for each calculation error

base: [2 × 1] [2 × 0]

1: wears seat belt

0: does not

number of draws: 500

sample % = $\frac{346}{500} = 0.692 = 69.2\%$ (4)

SD base = SD sample = $\sqrt{\frac{346}{500} \cdot \frac{154}{500}} = \sqrt{0.213} = 0.46$ (3)

$SE_{sum} = \sqrt{500} \cdot 0.46 = 22.36 \cdot 0.46 = 10.28$ (2)

$SE_{\%} = \frac{10.28}{500} \cdot 100\% = 2.06\%$ (2)

95% CI: $69.2\% \pm 2 \cdot 2.06\% = 69.2\% \pm 4.12\%$ (4)

$= 65.08\%$ to 73.32%

-2 for each incorrect value used here

2) (5 Points) Can we do the same calculation as in part 1) if it turns out that 495 out of 500 drivers always wear a seat belt? Yes or no. **Explain!**

No! On page 168 of the lecture notes, we state that "the sample percentage should not be too close to 0% or 100%". But in this case, the sample % would be $\frac{495}{500} = 99\%$, i.e., very close to 100%. So, this sample % would be too close to 100% to give a valid 95% CI. (The reason is that the upper bound of this CI almost exceeds 100%.)

(3): No

(2): Reasonable explanation

(1): Yes, some explanation

(0): Yes, no explanation

Please turn over!

Statistics 1040, Sections 002, 003 & 004, Quiz 11 (20 Points)

April 19, 2002

Your Name: _____

Question 1: Tests of Significance(**20 Points**)

National data show that on average, college freshmen spend 7.5 hours a week going to parties. One administrator does not believe that these figures apply to her college, which has nearly 3000 freshmen. She takes a simple random sample of 100 freshmen and interviews them. On average, they report 6.6 hours a week going to parties, and the SD is 9 hours. Does it mean that the students from this college really spend less time going to parties than the national average, or is it just a chance variation?

1. (5 points) State the null and the alternative hypothesis for this problem, in words and in terms of the box model.

Null: freshmen from this college spend 7.5 h a week going to parties, i.e., box avg = 7.5

Alternative: freshmen from this college spend less than 7.5 h a week going to parties, i.e., box avg < 7.5 h

2. (5 points) Calculate the appropriate test statistic.

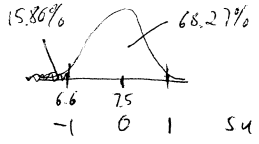
observed: 6.6 expected: 7.5

$SE_{sum} = \sqrt{100} \cdot 9 = 90$

$SE_{avg} = \frac{90}{100} = 0.9$

$\rightsquigarrow z = \frac{obs - exp}{SE} = \frac{6.6 - 7.5}{0.9} = -1.00$

3. (5 points) Obtain the P-value (use the normal table on the back).



P-value = 15.86%

4. (5 points) State conclusions in terms of rejecting the null hypothesis and in your own words.

P-value greater than 5%

→ do not reject null hypothesis

→ freshmen from this college spend the same amount of time on parties as freshmen at other colleges

Please turn over!

Grading Criteria:

1) wrong null and alternative hypotheses, e.g.:
- mixed null and alternative −3
- "6.6" in hypothesis instead of "7.5" −3

hypothesis stated in words only or in numbers only −1 each

2) incorrect z, e.g. $\frac{avg - obs}{SE}$ −3

incorrect SE_{avg} −2

calculation error −1 each

3) incorrect area −2

incorrect table value −2

calculation error −1

4) reject null if p-value > 5% −4

(or do not reject null if p-value < 5% −4)

no explanation (e.g., reject, but no conclusion) −2

Correctly rejecting, but explanation missed up −2

if not speaking of rejecting / not rejecting −2

Statistics 1040, Sections 002, 003 & 004, Quiz 12 (20+ Points)

April 26, 2002

Your Name: _____

You have exactly 15 minutes for this quiz. Each of the three questions is worth 20 points. Choose one question as your "main question" and do it completely. Clearly indicate which question you have chosen. You may also earn extra credit if you manage to work on more than one questions during this time.

Question 1: from: Stat 1040, Fall 2000, Final Test, Thursday, December 14, 2000 (Question 9)

(20 Points) The spermicide nonoxynol–9 kills HIV in the test tube, so researchers hypothesized that it might be useful in protecting high–risk women from HIV. Other researchers argued that nonoxynol–9 might increase the risk because it is an irritant. In a study of 990 prostitutes, participants were randomly divided into two groups. The treatment group were given a nonoxynol–9 gel. The control group were given a similar–looking but inactive gel. When the study ended in May 2000, 67 of the 495 women in the treatment group were HIV–positive, and 44 of the 495 women in the control group were HIV–positive. Perform a 2–tailed test to decide whether the treatment and control groups were significantly different with respect to HIV. Clearly state your conclusions.

2-sample z-test

① Null: treatment makes no difference w.r.t. HIV, i.e., $\%_{treat} = \%_{control}$

Alternative: treatment does make a difference, i.e., $\%_{treat} \neq \%_{control}$

② Treatment

$SD = \sqrt{\frac{67}{495} \cdot \frac{428}{495}} = 0.342$

$\%_{treat} = \frac{67}{495} = 13.5\%$

$SE_{sum\ treat} = \sqrt{495} \cdot 0.342 = 7.6$

$SE_{\%\ treat} = \frac{7.6}{495} \cdot 100\% = 1.5\%$

Control

$SD = \sqrt{\frac{44}{495} \cdot \frac{451}{495}} = 0.285$

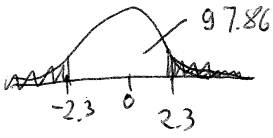
$\%_{control} = \frac{44}{495} = 8.9\%$

$SE_{sum\ control} = \sqrt{495} \cdot 0.285 = 6.3$

$SE_{\%\ control} = \frac{6.3}{495} \cdot 100\% = 1.3\%$

$SE_{diff} = \sqrt{1.5^2 + 1.3^2} = 2.0\%$

$z = \frac{13.5\% - 8.9\%}{2.0\%} = 2.3$

③ 

P-value = area outside = 100% − 97.86% = 2.14%

④ reject the null hypothesis (P-value < 5%)

- result statistically significant
- treatment makes a difference w.r.t. HIV (i.e., nonoxynol-9 increases the risk of HIV)

Question 2: from: Stat 1040, Spring 2000, Final Test, Friday, May 5, 2000 (Question 5)

(20 Points) Reading test scores are known to follow the normal curve. The average reading test score is supposed to be 100. I suspect that the children at one of the local schools have a higher average reading score, so I take a simple random sample of 10 of these children and find that their average reading score is 112, with an SD of 18. Is my suspicion correct? Perform the appropriate statistical test. You must clearly state a null and alternative hypothesis, compute a test statistic and a P-value and state your conclusions.

sample size = 10 (< 30), assume "normal" & SD box unknown → t-test

① Null: same score as elsewhere, i.e., box avg (for all children at this school) = 100
Alternative: higher score as elsewhere, i.e., box avg > 100

② $SD^+ = \sqrt{\frac{10}{10-1}} \cdot 18 = 19$

$SE_{sum} = \sqrt{10} \cdot 19 = 60$

$SE_{avg} = \frac{60}{10} = 6$

$t = \frac{112 - 100}{6} = 2$ $\quad df = 10 - 1 = 9$

③ in table: 2.26 ↝ 2.5%, 1.83 ↝ 5%
so t = 2 ↝ p-value between 2.5% and 5%

④ reject the null hypothesis (p-value < 5%)
- result statistically significant
- students at this school have a higher average score than students elsewhere

Question 3: from: Stat 1040, Spring 1999, Final Test, Monday, May 3, 1999 (Question 8)

(20 Points) In the game of chess, the first few moves play a very important role in determining the final outcome. Five different opening strategies are highly favored by chess experts. To determine whether one or more of these strategies is most preferred by grand masters in international competition, a random sample of 100 grand masters is taken, and each is asked which of the strategies he or she would prefer to employ. A summary of their responses is given below:

Strategy	A	B	C	D	E
Frequency	17	27	22	15	19

Make a χ^2–test of the null hypothesis that there is no preference between these strategies by grand masters in international competition. You should state the null and the alternative hypotheses and clearly state your conclusions.

① Null: no preference – choosing at random, chance for each strategy is 1/5
Alternative: some strategies are preferred, chances not all equal to 1/5

②

Strategy	Obs	Exp	$\frac{(Obs-Exp)^2}{Exp}$
A	17	20	0.45
B	27	20	2.45
C	22	20	0.20
D	15	20	1.25
E	19	20	0.05

sum → $\chi^2 = 4.4$

$df = 5 - 1 = 4$

③ 

p-value between 30% and 50%
($\chi^2 = 4.4$ between 4.88 and 3.36)

④ do not reject the null hypothesis (p-value > 5%)
- no preference – strategies chosen at random