Statistics 1040 Final Exam, Fall 2005

Name or A-Number:

100 → $\boxed{400}$

Directions: You have 110 minutes to complete the exam. The exam will be graded out of 100 points. Be sure to answer every question. You must show your work for full credit.

---

1. Read the following news article:

22 → $\boxed{88}$

"Wrap up" advice to stop colds

Scientists say cold noses reduce ability to fight virus attacks

Monday, November 14, 2005; Posted: 4:47 p.m. EST (21:47 GMT)

LONDON, England (CNN) – British researchers into the common cold say "catching a chill" really does help colds develop – and are advising to "wrap up warm" to keep viruses at bay.

Mothers and grandmothers have long warned that chilling the surface of the body, through wet clothes, feet and hair, causes common cold symptoms to develop. But much previous research has dismissed any link between chilling and viral infection as having no scientific basis.

Now researchers in Cardiff, Wales, say they can prove drops in temperature to the body really can cause a cold to develop.

Claire Johnson and Professor Ron Eccles, from Cardiff University's Common Cold Center, recruited 180 volunteers, half of whom they got to immerse their feet in ice and cold water for 20 minutes.

The other 90 in tests during the common cold "season" sat with their feet in an empty bowl.

During the next four or five days, almost a third (29 percent) of the chilled volunteers developed cold symptoms – compared to just 9 percent in the control group, the scientists said.

$\boxed{8}$ (a) (2 points) Is this an observational study or a designed experiment? Explain briefly. 6

Participants were told 2 what to do (hold feet in ice or in an empty bowl).

$\boxed{8}$ (b) (2 points) What is the "treatment" in this study?

Holding feet in ice and cold water for 20 minutes. 8

$\boxed{12}$ (c) (3 points) The article does not say how the researchers divided up the 180 people into two groups of 90. How *should* they do this? Explain clearly.

28

Participants should be assigned randomly 6 to the two groups, perhaps in this way: Number participants from 1 to 180 and blindly draw 90 numbers and place participants with these numbers in the treatment group. Place the remaining 90 participants in the control group. 6

8 (d) (2 points) Is the study blind? Explain briefly.

It is not 4 blind. Participants 4 knew whether they had their feet in ice or in an empty bowl.

12 (e) (3 points) Are there any possible confounding factors or problems with the study? Explain clearly.

20

First, the study is not blind. Participants who had their feet in ice might tend to report a cold more frequently than participants who had their feet in an empty bowl.

any valid factor 8

explanation 4

Then, there are several possible confounding factors:

- Age: Older people and little children might develop a cold more easily than medium-aged people.
- Overall Health Status: People with other health problems might develop a cold more easily than healthier people.
- Job: People working around sick people (hospital or doctor's office) or around many other people (teachers) might develop a cold more easily.

40 (f) (10 points) Assume that the researchers followed your instructions in part (c) and ignore any other problems of which you might be aware. Using the numbers provided in the article, perform a statistical hypothesis test that the researchers could have used to test their belief that drops in temperature can cause a cold to develop.

You must clearly state the null and the alternative hypotheses, compute a test statistic and a P-value and clearly state your conclusions in terms of the language used in the newspaper article.

I: People exposed to ice

E: People exposed to an empty bowl

-2 for each calculation error

-30 for incorrect test

-4 if null, alt swapped

2-sample z-test:

1) null: I and E develop cold at the same rate, 2

i.e., $box_{I\%} - box_{E\%} = 0\%$ 1

alternative: I develop cold at a higher rate, 2

i.e., $box_{I\%} - box_{E\%} > 0\%$ 1

2)

I	E
sample size I = 90	sample size E = 90
sample I% = 29%	sample E% = 9%

$SD_I = \sqrt{0.29 \cdot 0.71} = 0.45$ 2 $SD_E = \sqrt{0.09 \cdot 0.91} = 0.29$ 2

$SE\,sum_I = \sqrt{90} \cdot 0.45 = 4.27$ 2 $SE\,sum_E = \sqrt{90} \cdot 0.29 = 2.75$ 2

$SE\%_I = \frac{4.27}{90} \cdot 100\% = 4.74\%$ 2 $SE\%_E = \frac{2.75}{90} \cdot 100\% = 3.06\%$ 2

$SE_{diff\%} = \sqrt{(4.74\%)^2 + (3.06\%)^2} = 5.64\%$ 4

$z = \frac{29\% - 9\%}{5.64\%} = 3.55$ 3

3) area between -3.55 and 3.55: 99.961% 3

$\rightsquigarrow$ P-value: $\frac{100\% - 99.961\%}{2} = 0.0195\%$ 3

$\approx 0.02\%$

4) • reject the null (P-value < 5%) 3

• result is highly stat. sig. (P-value < 1%) 3

• people exposed to ice develop cold at a higher rate than people exposed to an empty bowl 3

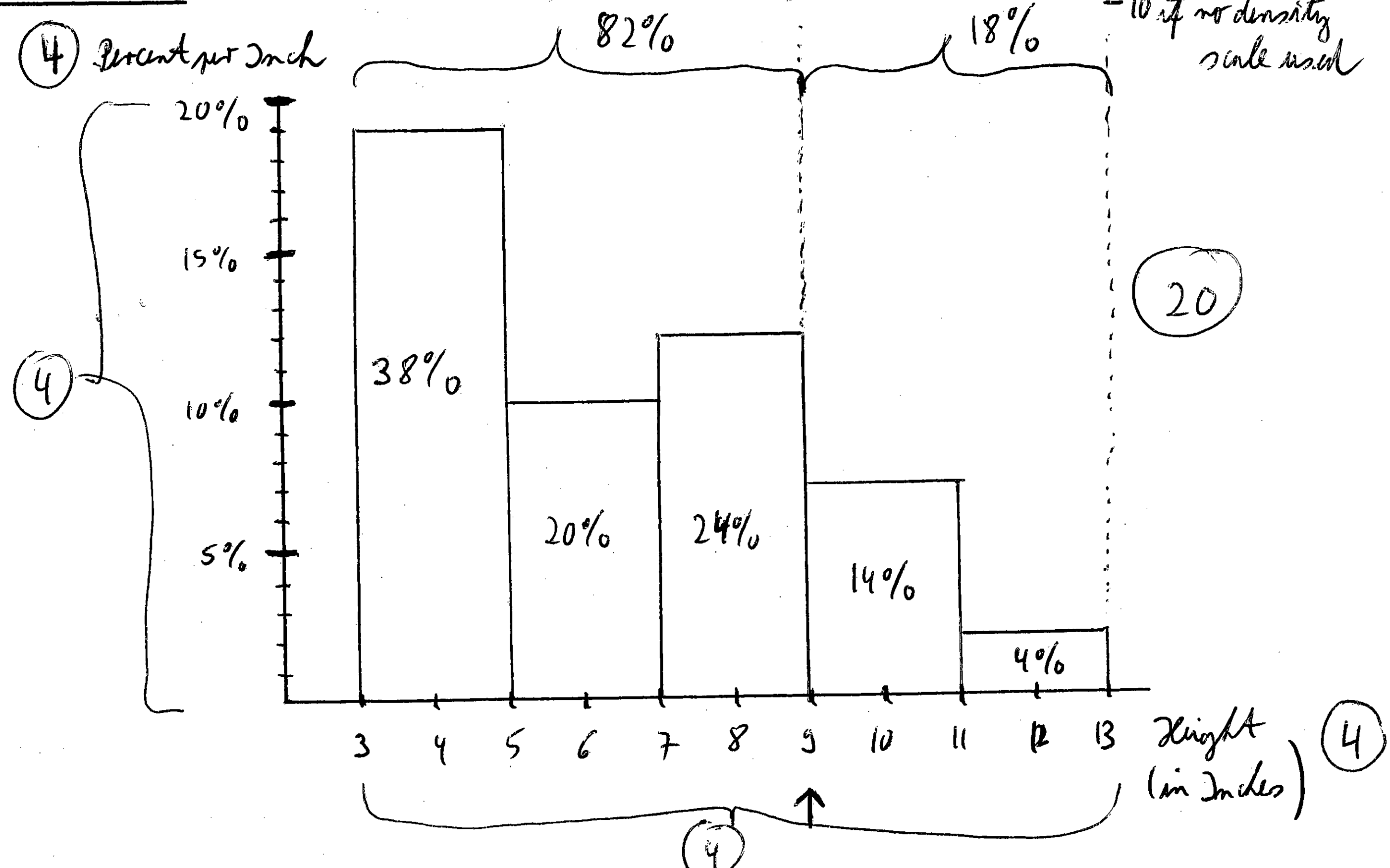
11 → $\boxed{44}$

2. When the Tribbles invaded the spaceship Enterprise, suppose that crew member Spock decided to take the logical step of seeing what the crew was up against, and he wanted to graphically represent the sizes of the Tribbles. Suppose that the table below summarizes the heights of the 50 Tribbles he found on the bridge. (Class intervals include the left but not the right endpoints.)

Tribble Height (inches)	Number of Tribbles	Percentage	Width	Height = Percentage / Width
3-5	19	19/50 = 38%	2	38% /2 = 19%
5-7	10	10/50 = 20%	2	20% /2 = 10%
7-9	12	12/50 = 24%	2	24% /2 = 12%
9-11	7	7/50 = 14%	2	14% /2 = 7%
11-13	2	2/50 = 4%	2	4% /2 = 2%
	50	100%		

$\boxed{36}$ a. (9 points) Draw a histogram of these height data, with the vertical axis on the usual density scale, both axes labeled, and heights of bars clearly indicated.

−10 if no density scale used



$\boxed{8}$

b. (2 points) If a Tribble is in the 82^{nd} percentile for height, about how tall is it? (Note: Use the histogram, NOT the normal curve).

82nd percentile means: 82% to the left of this value and 8% to the right of this value; this is at a height of approx. 9 inches. 8

$7 \rightarrow \boxed{28}$

3. For college-aged men, the average height is 70 inches with a standard deviation of 3 inches, and the average weight is 162 pounds with a standard deviation of 30 pounds. The correlation between height and weight for college-aged men is 0.47.

$\boxed{12}$ a. (3 points) Find the equation of the line to predict weight from height for college-aged men.

-1 if x,y not stated
-3 if x,y swapped
-2 for each calculation error

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.47 \cdot \frac{30}{3} = 4.7$$

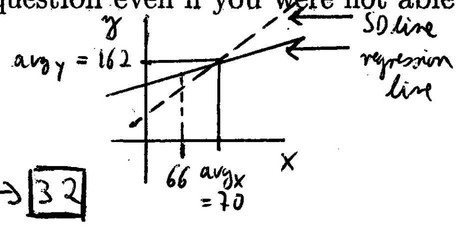
$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 162 - 4.7 \cdot 70 = -167$$

regression equation: $\boxed{y = -167 + 4.7 \cdot x}$ or $\boxed{\text{weight} = -167 + 4.7 \cdot \text{height}}$ 3

$\boxed{8}$ b. (2 points) Predict the weight for a college-aged man who is 66 inches tall.

for height 66: weight $= -167 + 4.7 \cdot 66 = 143.2$ pounds 8

$\boxed{8}$ c. (2 points) If the man in part b is exactly on the SD line, will his true weight be larger or smaller than your answer to part b? Explain briefly. (Note that it is possible to answer this question even if you were not able to answer part b.) 4



smaller! The SD line is steeper than the regression line; we are looking at a person with a weight below the average, so the SD line is below the regression line on this side. 4

Note: $\frac{66-70}{3} = -1.33$ s.u. below avg_x; to be on the SD line, this person must have a weight of $162 - 1.33 \cdot 30 = 122$ pounds – which is smaller than 143.2 pounds.

$8 \rightarrow \boxed{32}$

4. In one Stat 1040 class, there are 48 students, of whom 13 are male and 35 are female. Of the male students, 6 are from the College of Education, and of the female students, 16 are from the College of Education. Two students are selected at random (like drawing without replacement) from the class.

-2 for each calculation error

$\boxed{8}$ a. (2 points) What is the chance that both students are women?

1st woman: 35/48
2nd woman (given 1st woman): 34/47

both women: $35/48 \cdot 34/47 = 0.527 = 52.7\%$

$\boxed{8}$ b. (2 points) What is the chance that neither of the students is male? 8

neither male = 1st not male & 2nd not male = 1st woman & 2nd woman = 52.7% as in part a)

$\boxed{8}$ c. (2 points) What is the chance that the first student is male and the second student is female?

1st male: 13/48
2nd female (given 1st male): 35/47

1st male & 2nd female: $13/48 \cdot 35/47 = 0.202 = 20.2\%$

$\boxed{8}$ d. (2 points) What is the chance that at least one of the students is from the College of Education? =22

6+16 students from Education & 48 − 22 = 26 students not from Education

1st not from Ed: 26/48
2nd not from Ed (given 1st not from Ed): 25/47

both not from Ed: $26/48 \cdot 25/47 = 0.288$
at least 1 from Ed: $1 - 0.288 = 0.712 = 71.2\%$

4 → $\boxed{16}$

5. (4 points) A gambler loses ten times running at a game of chance. The gambler thinks he should keep playing because he is due for a win by the law of averages. A bystander advises him to quit, on the grounds that his luck is cold. What does statistics say?

Statistics says: In games of chance (such as flipping coins or rolling dice), outcomes of one play are <u>independent</u> (8) of previous outcomes. Neither the law of averages nor luck should be cited here – the chances to win remain the same for the next play as they were in all previous plays! (8) for overall explanation

Note: Nico Zographos, casino dealer, said in the 1920s: "There is no such thing as luck. It is all Mathematics." [see Chance, Vol. 18, No. 2, 2005, p. 41]

12 → $\boxed{48}$

6. Suppose researchers selected a simple random sample of 1,200 U.S. taxpayers and found that in 2004 these 1,200 received an average tax refund of $2,052, with a standard deviation of $431. According to the IRS, the average refund for all U.S. taxpayers that year was $2,063.

$\boxed{8}$ a. (2 points) What does it mean when we say that the researchers selected a "simple random sample" of 1,200 U.S. taxpayers?

- they may have numbered all US taxpayers and then have drawn 1200 numbers <u>randomly</u> (4) and included those taxpayers represented by the numbers they have drawn
- this implies that 1) all US taxpayers have the same chance to be in the sample (2) & 2) each possible combination of 1200 US taxpayers has the same chance of being selected (2)

b. (7 points) Using the researchers' results, find a 90 percent confidence interval for the average refund for all U.S. taxpayers in 2004.

$\boxed{28}$

−2 for each calculation error

avg = 2,052

SD = 431

$SE_{sum} = \sqrt{1{,}200} \cdot 431 = 14{,}930$ (4)

$SE_{avg} = \dfrac{14{,}930}{1{,}200} = 12.4$ (4)

90% CI: $2{,}052 \pm 1.65 \cdot 12.4 = 2{,}052 \pm 20.5 = 2031.5$ to 2072.5

(8) (4) (2) (4)(2)(4) (2) (2)

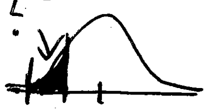
$\boxed{12}$ c. (3 points) True or False and explain briefly: The sizes of the tax returns of 90% of all U.S. taxpayers in 2004 are in this 90 percent confidence interval.

We are 90% confident that the true population average tax return is in this interval. This does <u>not</u> mean that 90% of all US taxpayers tax returns in 2004 fall into this interval. (4)

$6 \to \boxed{24}$ — 2 for each calculation error

7. (6 points) Suppose that 13 percent of all people are left-handed. If a simple random sample of seven hundred people is considered, what are the chances that between 9.3 percent and 11.2 percent of them are left-handed?

2



box: $\boxed{13 \times \boxed{1} \quad 87 \times \boxed{0}}$ (2) 1: left-handed, 0: right-handed

#draws: 700 (2)

box avg = 0.13

box SD = $\sqrt{0.13 \cdot 0.87} = 0.34$ (2)

EV% = 13% (2)

$SE_{sum} = \sqrt{700} \cdot 0.34 = 9.00$ (2)

$SE\% = \frac{9.00}{700} \cdot 100\% = 1.29\%$ (2)

s.u.: $\frac{9.3\% - 13\%}{1.29\%} = -2.87$ (2)

$\frac{11.2\% - 13\%}{1.29\%} = -1.40$ (2)

area between −2.85 and 2.85: 99.56% (2)

area between −1.40 and 1.40: 83.85% (2)

area between −2.85 and −1.40: $\frac{99.56\% - 83.85\%}{2} = 7.86\%$ (4)

$10 \to \boxed{40}$

8. (10 points) On the M&M web page, it claims that they produce 13% brown, 14% yellow, 13% red, 24% blue, 20% orange, and 16% green milk chocolate M&M's. Suppose we buy a bag of milk chocolate M&M's and come up with the following numbers of each color:

−2 for each calculation error
−30 for incorrect test
−4 if null, alt swapped

Color	Number (= obs count)	expected %	expected count	$\frac{(obs-exp)^2}{exp}$
brown	50	13%	381 · .13 = 49.5	$\frac{(50-49.5)^2}{49.5} = 0.005$
yellow	47	14%	381 · .14 = 53.3	$\frac{(47-53.3)^2}{53.3} = 0.745$
red	41	13%	381 · .13 = 49.5	1.460
blue	94	24%	381 · .24 = 91.4	0.074
orange	102	20%	381 · .20 = 76.2	8.735
green	47	16%	381 · .16 = 61.0	3.213
	381	100%	380.9 (12)	$\chi^2 = 14.232$ (4)

Test the hypothesis that our bag of M&M's is like a simple random sample of M&M's from a population with the specified percentages of each color. You must state a null and an alternative hypothesis, find a test statistic and a P-value and clearly state your conclusions in terms of what you have learned about the color of milk chocolate M&M's.

χ^2-test for distribution:

1, null: mix is OK (3), i.e., percentages are 13% brown, 14% yellow, 13% red, 24% blue, 20% orange and 16% green (1)

alternative: mix is not OK (3), i.e., at least one of the percentages is different from the claims on the web page (1)

2| see table above for some of the calculations.

$\chi^2 = \text{sum of } \frac{(obs-exp)^2}{exp} = 14.232$

$df = 6 - 1 = 5$ (3)

3) $\chi^2 = 14.232$ is between 11.07 and 15.09 (2)

$\rightsquigarrow$ P-value is between 5% and 1% (2)

4) • reject the null (P-value < 5%) (3)

• result is stat. sig. (P-value > 1%) (3)

• mix is not OK; at least one of the percentages is different from the claims on the web page (3)

10 -> $\boxed{40}$

9. A Stat 1040 instructor recently tried to answer the following question:

> "Does it help to improve Stat 1040 students' quiz scores when the answers to a quiz are handed out the lecture *before* the actual quiz?"

On Monday, students were given a handout with some worked examples. The students were strongly advised to look through this handout to prepare for the Wednesday quiz. The quiz consisted of one of the problems from the handout. The 38 participating students averaged 13.5 points (out of 20) with an SD of 5.8 points. In the past, students of this instructor averaged 11.4 points (out of 20) on a similar quiz. Do his Stat 1040 students score higher if they have access to the solutions before the quiz, or not? For the purpose of this question, you should assume that the 38 students who took this particular quiz are like a simple random sample of all Stat 1040 students taught by this instructor.

−2 for each calculation error
−30 for incorrect test
−4 if null, alt swapped

$\boxed{8}$ (a) (2 points) Clearly state the null and the alternative hypotheses.

z-test: sample size > 30

1) null: students performed as in the past, i.e., box$_{avg}$ = 11.4 (1) (3)

alternative: students performed better than in the past, (3) i.e, box$_{avg}$ > 11.4 (1)

$\boxed{20}$ (b) (5 points) Compute a test statistic.

2) obs avg = 13.5

exp avg = 11.4

$SE_{sum} = \sqrt{38} \cdot 5.8 = 35.75$ (6)

$SE_{avg} = \frac{35.75}{38} = 0.94$ (6)

$z = \frac{13.5 - 11.4}{0.94} = 2.23$ (8)

$\boxed{4}$ (c) (1 point) Find the P-value.

3) area between −2.25 and 2.25: 97.56% (2)

area above 2.25: $\frac{100\% - 97.56\%}{2}$ = 1.22% = P-value (2)

$\boxed{4}$ (d) (1 point) Do you reject the null hypothesis? Explain why or why not.

4) • reject the null (2) (P-value < 5%) (2)

$\boxed{4}$ (e) (1 point) Clearly state your conclusions.

• result is stat sig. (P-value > 1%) (2)

• it appears that students performed better than in the past (2)

(but, hey, why didn't they average near 20 points when they got the solutions before the actual quiz ?!?)

$10 \to \boxed{40}$

10. In a large city, there are 5 electoral precincts. There are two mayoral candidates, A and B. A political science student takes a simple random sample of 1870 voters from this city and asks them which precinct they live in and whether they voted for candidate A or B. She makes the following table:

obs count

Precinct	Candidate A	Candidate B	Total
1	400	229	629
2	184	154	338
3	101	200	301
4	182	186	368
5	136	98	234
Total	1003	867	1870

expected count:

337	292	629
181	157	338
161	140	301
197	171	368
126	108	234
1002	868	1870

−2 for each calculation error
−30 for incorrect test
−4 if null, alt swapped

$10 \times 1 = 10$

We want to test the hypothesis that voting for mayoral candidates A and B is independent of precinct for voters from this city.

$\boxed{8}$ (a) (2 points) Clearly state the null and the alternative hypotheses.

χ^2-test for independence:

1/ null: voting and precinct are independent, (3) i.e., boxes are identical (1)

alternative: voting and precinct are not independent (voting changes with precinct), (3) i.e., at least one box is different (1)

$\boxed{16}$ (b) (4 points) Compute a test statistic.

2/ expected: $\frac{629 \cdot 1003}{1870} = 337$ etc. (see table "expected count" above)

$$\chi^2 = \text{sum of } \frac{(\text{obs}-\text{exp})^2}{\text{exp}} = \frac{(400-337)^2}{337} + \frac{(184-181)^2}{181} + \frac{(101-161)^2}{161} + \frac{(182-197)^2}{197} + \frac{(136-126)^2}{126}$$

$$+ \frac{(229-292)^2}{292} + \frac{(154-157)^2}{157} + \frac{(200-140)^2}{140} + \frac{(186-171)^2}{171} + \frac{(98-108)^2}{108}$$

$$= 77.73 \quad (6)$$

$\boxed{4}$ (c) (1 point) Find the degrees of freedom.

$df = (5-1)\cdot(2-1) = 4$ (4)

$\boxed{4}$ (d) (1 point) What can you say about the P-value?

3/ $\chi^2 = 77.73$ is to the right of 13.28 (2)

⇝ P-value is to the right of 1%, i.e., P-value < 1% (2)

$\boxed{4}$ (e) (1 point) Do you reject the null hypothesis? Explain why or why not.

4/ • reject the null (2) (P-value < 5%) (2)

$\boxed{4}$ (f) (1 point) Clearly state your conclusions.

- result is highly stat. sig. (P-value < 1%) (2)
- voting and precinct are not independent (voting changes with precinct) (2)