Stat 1040, Fall 2007 **Name:** ____________________

Final Test, December 13, 11:30pm–1:20pm

100 → $\boxed{400}$

Show your work. The test is out of 100 points and you have 110 minutes to finish.

7 → $\boxed{28}$ 1. In the December 10 issue of NEWSWEEK medical writer Jerry Adler says:

> It's not too soon to start thinking about New Year's resolutions, and here's mine, as a medical writer: I will not report on any amazing new treatments for anything, unless they were tested in large, randomized, placebo-controlled, double-blind clinical trials published in high-quality peer-reviewed medical journals. If that means not telling NEWSWEEK's readers about, say, a new magnetized-water cure for osteoporosis, cancer and autism – well, there are infomercials to fill that gap.

$\boxed{8}$ (a) (2 points) Explain what it means for a study to be double-blind.

The subjects do not know whether they are in the treatment or in the control group; 4 nor do the doctors and nurses know who work with these subjects. 4

$\boxed{12}$ (b) (3 points) Give 3 different reasons *why* a medical study should be double-blind.

- it guards against bias in the subjects' responses 4
- it guards against bias in the doctors' and nurses' behavior towards the subjects 4
- it guards against bias in doctors' assessment of a disease, i.e., did the patient improve / fully recover from the disease or not? 4

$\boxed{8}$ (c) (2 points) What is a placebo? Why is it used?

- a placebo is a drug or vaccination (e.g., a sugar pill or a salt water injection) that resembles the treatment, but has no medical effect 4
- it is used such that the subjects' response will be related to the treatment itself and not to the idea of the treatment 4

5 → 20

2. (5 points) Psychologist Daniel Kahneman was teaching flight instructors that praise is more effective than punishment for promoting learning, when one of the most seasoned instructors in his audience raised his hand and said, "On many occasions I have praised flight cadets for clean execution of some aerobatic maneuver, and in general when they try it again, they do worse. On the other hand, I have often screamed at cadets for bad execution, and in general they do better the next time. So please don't tell us that reinforcement works and punishment does not, because the opposite is the case."

What does statistics say about this instructor's experience?

This is the regression effect! (8) In virtually all test-retest situations, the bottom group on the first test will on average show some improvement on the second test - and the top group will on average fall back. Thinking that the regression effect must be due to something important, and not just due to the spread of the data around the regression line, is the regression fallacy. (6)

(6) for general explanation

5 → 20

3. (5 points) For a sample of 570 California women age 25 to 29 in 2005 the relationship between education (years of schooling completed) and income can be summarized as follows:

x Average education ≈ 13.0 years, SD ≈ 3.4 years

y Average income ≈ $18,000, SD ≈ $20,000 r ≈ 0.37

− 2 for each calculation error

Predict the income for one of these women with 15 years of education.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.37 \cdot \frac{20{,}000}{3.4} \overset{(7)}{=} 2{,}176.5$$

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 18{,}000 - 2{,}176.5 \cdot 13 \overset{(7)}{=} -10{,}294.5$$

regression equation: $y = -10{,}294.5 + 2{,}176.5 \cdot x$

income for 15 years of education: $y = -10{,}294.5 + 2{,}176.5 \cdot 15 \overset{(6)}{=} \underline{\underline{\$22{,}353}}$

5 → 20

4. (5 points) A 1999 study claimed that

> Infants who sleep at night in a bedroom with a light on may be at higher risk for myopia (nearsightedness) later in childhood.

The researchers surveyed parents of 479 children aged 2 to 16 seen in the ophthalmology outpatient department of a children's hospital. A questionnaire asked about the child's nighttime light exposure at the time of the survey and before age two. They noticed a positive association between myopia and nighttime light exposure.

Explain why this is *not* strong evidence that sleeping with a light on *causes* myopia by suggesting a possible confounding factor and explaining how this confounding factor could account for the association they observed.

First of all, association is not causation! (10)

Possible confounding factors are:

- genetics: actually, nearsightedness tends to run in families
- moreover, http://en.wikipedia.org/wiki/Myopia lists ethnicity, race, education, intelligence and IQ as additional confounding factors
- age: a child aged 3 may not 2 have had the time to develop myopia, whereas a child aged 16 may have developed myopia for numerous reasons over time

(10) for one factor/explanation

8 → $\boxed{32}$

5. For a road trip, a student places the following ten CDs into the glove compartment of his car

- 6 *modern rock* CDs (Fallout Boy, Hawthorne Heights, The Used, Finger Eleven, Taking Back Sunday, She Wants Revenge),
- 3 *pop* CDs (P!nk, Fergie, Gwen Stefani),
- 1 *American Idol* CD (Jordin Sparks).

On his trip, the student blindly grabs a CD from the glove compartment, listens to it, and places it on the back seat when finished. Then he blindly grabs a second CD from the glove compartment. You should NOT comment on the musical taste of this student, but answer each of the following questions separately.

−2 for each calculation error

$\boxed{8}$ (a) (2 points) What is the chance that the SECOND CD will be a *pop* CD or the *American Idol* CD?

2nd pop, or (mutually exclusive), 2nd Idol

$\frac{3}{10}$ (3) + (2) $\frac{1}{10}$ (3) $= \frac{4}{10} = \underline{0.4} = 40\%$

$\boxed{8}$ (b) (2 points) What is the chance that he will listen to Jordin Sparks as one of his two selections?

1st Jordan and 2nd other or (mutually exclusive) 1st other and 2nd Jordin = 1st Jordin or (mutually exclusive) 2nd Jordin

$\frac{1}{10} \cdot \frac{9}{9}$ (3) + (2) $\frac{9}{10} \cdot \frac{1}{9}$ (3) $= \frac{1}{10} + \frac{1}{10} = \frac{2}{10} = \underline{0.2} = 20\%$

$\boxed{8}$ (c) (2 points) What is the chance that he will listen to none of the *pop* CDs?

1st not pop and 2nd not pop

$\frac{7}{10}$ (3) $\cdot$ (2) $\frac{6}{9}$ (3) $= \frac{42}{90} = \underline{0.467} = 46.7\%$

$\boxed{8}$ (d) (2 points) What is the chance that he will listen to at least one of the *modern rock* CDs?

opposite rule, 1st not modern rock and 2nd not modern rock

$1 -$ (2) $\frac{4}{10}$ (2) $\cdot$ (2) $\frac{3}{9}$ (2) $= \frac{90}{90} - \frac{12}{90} = \frac{78}{90} = \underline{0.867} = 86.7\%$

8 → $\boxed{32}$

6. (8 points) In October 29 to November 1 2007, a local survey organization questioned 603 residents in Utah. They found that 23% of those surveyed "definitely favor" building a nuclear power plant in Utah. Assuming this is a simple random sample of Utahns, find a 90% confidence interval for the percentage of all Utahns who would say that they "definitely favor" building a nuclear power plant in Utah.

−2 for each calculation error

sample % = 23%

$SD = \sqrt{0.23 \cdot 0.77} = 0.421$ (4)

$SE_{sum} = \sqrt{603} \cdot 0.421 = 10.34$ (4)

$SE_{\%} = \frac{10.34}{603} \cdot 100\% = 1.7\%$ (4)

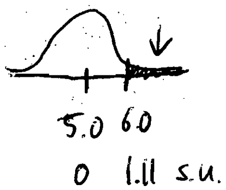
90% CI: 23% (4) $\pm$ (2) 1.65 (4) $\cdot$ (2) 1.7% (4) $= \underline{20.2\%}$ (2) to $\underline{25.8\%}$ (2)

33 → $\boxed{132}$

7. A grocery store carries a variety of "on the vine" tomatoes with an average weight of 5.0 ounces and an SD of 0.9 ounces. The weights of these tomatoes follow the normal curve.

$\boxed{24}$ (a) (6 points) What percentage of them would weigh more than 6.0 ounces?

−2 for each calculation error (in any part)

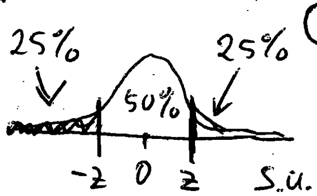


s.u.: $\frac{6.0-5.0}{0.9} = \frac{1.0}{0.9} = 1.11$ 8

area between −1.10 and 1.10: 72.87% 8

area above 1.10: $\frac{100\% - 72.87\%}{2} = \underline{\underline{13.57\%}}$ 8

$\boxed{24}$ (b) (6 points) Estimate the 25th percentile of their weights.



8

area between $\underline{\underline{-0.65}}$ and 0.65: 48.43% (closest to 50%)

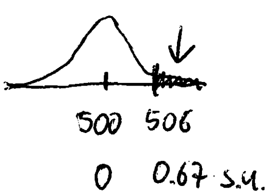
original units: $-0.65 \cdot 0.9 + 5.0 = \underline{\underline{4.42 \text{ ounces}}}$

8 2 2 2 2

$\boxed{24}$ (c) (6 points) Find the chance that the total weight of 100 randomly selected tomatoes will be more than 506 ounces.

$EV_{sum} = 100 \cdot 5.0 = 500$ 5

$SE_{sum} = \sqrt{100} \cdot 0.9 = 9$ 5



s.u.: $\frac{506-500}{9} = \frac{6}{9} = 0.67$ 5

area between −0.65 and 0.65: 48.43% 5

area above 0.65: $\frac{100\% - 48.43\%}{2} = \underline{\underline{25.79\%}}$ 4

Note: Part (d) should be a z-test as the box SD is given on the previous page! However, this is so easy to miss that a t-test comes to mind at first. In this exam, you get full points for a z-test or a t-test.

[40] (d) (10 points) I'm a little skeptical of the claim that the average weight of the tomatoes is 5.0 ounces - I think it might be somewhat greater than 5.0 ounces. I select 9 tomatoes at random and find the following weights:

−30 for incorrect test (other than z or t)

−4 if null, alt swapped

6.1 4.4 6.3 5.7 4.5 6.9 5.1 5.8 5.7

Is there evidence that the average weight of this type of tomato is greater than 5.0 ounces? Clearly state the null and alternative hypotheses, calculate the appropriate test statistic, find the P-value, and state your conclusion.

at first glance

t-test:
- sample size < 30 ✓
- SD for box unknown ✓
- data follow normal curve ✓

1) null: tomatoes have correct weight, (2) i.e., box avg = 5.0 (1)

alternative: tomatoes are heavier, (2) i.e., box avg > 5.0 (1)

2) observed (avg) = $\frac{6.1+4.4+...+5.7}{9} = 5.6$ (3)

expected (avg) = 5.0

$SD = \sqrt{\frac{(6.1-5.6)^2+(4.4-5.6)^2+...+(5.7-5.6)^2}{9}} = 0.77$ (3)

$SD^+ = \sqrt{\frac{9}{8}} \cdot 0.77 = 0.82$ (3)

$SE_{sum} = \sqrt{9} \cdot 0.82 = 2.46$ (2)

$SE_{avg} = \frac{2.46}{9} = 0.27$ (2)

$t = \frac{5.6-5.0}{0.27} = 2.22$ (3)

3) df = 9 − 1 = 8 (3)

t = 2.22 between 1.86 and 2.31 (3)

P-value between 5% and 2.5% (3)

4)
- reject the null (3)
- result is stat. significant (3)
- tomatoes are heavier (3)

[20] (e) (5 points) Which, if any, of your answers to the first 4 parts of this question would still be reliable if you were told that the weights of the tomatoes did NOT follow the normal curve? Explain.

(d) at second glance

z-test:
- sample size < 30 ✓
- SD for box known (0.9) ✓
- data follow normal curve

1) as above (6)

different parts (no SD & SD+ calculations): (3) observed (avg) (4)

2) $SE_{sum} = \sqrt{9} \cdot 0.9 = 2.7$ (4)

$SE_{avg} = \frac{2.7}{9} = 0.30$ (4)

$z = \frac{5.6-5.0}{0.30} = 2.0$ (5)

3) area between −2.0 and 2.0: 95.45% (5)

$P\text{-value} = \frac{100\% - 95.45\%}{2} = 2.28\%$ (5)

4) as above (9)

(e)

- parts (a), (b), (d) [both for the t-test and the z-test] require normality
- (c) does not require normality and therefore would still be reliable; (10) recall the Central Limit Theorem (5) (FPP, p. 325): When drawing at random with replacement from a box, the probability histogram for the sum will follow the normal curve, even if the contents of the box do not. (5)

−8 for any additionally listed part

① $= \frac{4346 \cdot 3862}{7963} = 2108$ ② $= 4346 - 2108 = 2238$

③ $= \frac{1996 \cdot 3862}{7963} = 968$ ④ $= 1996 - 968 = 1028$

⑤ $= 3862 - 968 - 2108 = 786$ ⑥ $= 1621 - 786 = 835$

11 → $\boxed{44}$ χ^2-test for independence

8. A recent "live vote" survey on MSNBC.com asked participants whether men or women talked most. Anyone visiting the site was allowed to vote as many times as they wanted. The results are summarized in the following table.

−27 for incorrect test
−4 if null, alt swapped
−2 for each calculation error

obs. count

Response	Gender of the participant: Female	Male	*Total*
Women talk more	1352	2994	4346
Men talk more	1545	451	1996
Evenly split	965	656	1621
Total	3862	4101	7963

exp. count

① 2108	② 2238	4346
③ 968	④ 1028	1996
⑤ 786	⑥ 835	1621
3862	4101	7963

6 × ① = ⑥

For parts (a) through (g), treat this as a simple random sample from a population of interest, and suppose we are interested in knowing whether a person's gender is independent of their response for this population.

1) $\boxed{4}$ (a) (1 point) Clearly state the null hypothesis.

null: gender and response are independent, ③ i.e., boxes are identical ①

$\boxed{4}$ (b) (1 point) Clearly state the alternative hypothesis.

alternative: gender and response are <u>not</u> independent, ③ i.e., at least one box is different ①

2) $\boxed{12}$ (c) (3 points) Find the appropriate test statistic. (Note: if you cannot calculate the answer to this part, use test statistic = 10. This is not the correct answer, but you can them proceed with the rest of the problem).

expected: → see above

$$\chi^2 = \text{sum of } \frac{(\text{obs}-\text{exp})^2}{\text{exp}} = \frac{(1352-2108)^2}{2108} + \frac{(2994-2238)^2}{2238} + \frac{(1545-968)^2}{968} + \frac{(451-1028)^2}{1028} + \frac{(965-786)^2}{786} + \frac{(656-835)^2}{835} = 1273.4$$ ⑥

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

$df = (3-1)\cdot(2-1) = 2$ ④

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

$\chi^2 = 1273.4$ (far) to the right of 9.21 ② ⇝ P-value (much) smaller than 1% ②

4) $\boxed{4}$ (f) (1 points) Do you reject the null hypothesis? **Why or why not?**

- yes, reject the null ②
- result is highly stat. significant, i.e. the P-value is (much) smaller than 1% ②

$\boxed{4}$ (g) (1 points) For this population, what are your conclusions about a person's gender and their response?

- gender and response are <u>not</u> independent ④

$\boxed{8}$ (h) (2 points) On the web page, MSNBC.com states that this is "not a scientific survey" and points to a page titled "How 1,000 people can be more representative than 200,000". Give 2 *different* reasons why this survey would not be representative of the population of people who visit that particular website.

④ for each valid reason

- the survey is biased towards people who have a computer and access this particular web site (and this is <u>not</u> a simple random sample)
- such surveys often are answered by people with strong opinions
- people are allowed to answer more than once

8→$\boxed{32}$

9. (8 points) National data show that the number of years of schooling of people age 18 and over has an average of 13 years. A simple random sample of 700 people age 18 and over from a certain county has an average of 14 years of schooling, with an SD of 5 years. Can the difference between the average for the nation and the average for the sample be due to chance error or is there evidence that this county is different from the nation? Clearly state the null and alternative hypotheses, calculate the appropriate test statistic, find the P-value, and state your conclusion.

−24 for incorrect test
−4 if null, alt swapped
−2 for each calculation error

z-test: • sample size > 30

1, null: schooling in this county is as in the nation, (2)
i.e., box avg = 13 (1)
alternative: schooling in this county is different from the nation, (2)
i.e., box avg ≠ 13 (1)

2, observed (avg) = 14
expected (avg) = 13
SD = 5

$SE_{sum} = \sqrt{700} \cdot 5 = 132.3$ (3)

$SE_{avg} = \frac{132.3}{700} = 0.19$ (3)

$z = \frac{14-13}{0.19} = 5.26$ (3)

3, area between −5.26 and 5.26: almost 100% (4)
P-value = area outside = almost 0% (4)

4, • reject the null (3)
• result is highly stat. significant (3)
• schooling in this county is different from the nation (3)

10→$\boxed{40}$

10. (10 points) In a randomized, controlled, double-blind study published in The *Journal of the American Medical Association* in October 2007, researchers followed 371 heavy drinkers for 14 weeks to try to determine whether the migraine drug Topamax could help them to quit drinking. By the end of the study, 27 of the 183 people in the Topamax group had quit drinking completely, while only 6 of the 188 people in the placebo group had quit drinking completely. Is this evidence that Topamax helps, or could the result just be due to chance error? Clearly state the null and alternative hypotheses, calculate the appropriate test statistic, find the P-value, and state your conclusion.

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

T: Topomax group
C: Control group (placebo)

2-sample z-test:

1, null: T and C help quit drinking at the same rate, (2)
i.e., $box_{T\%} - box_{C\%} = 0\%$ (1)
alternative: T helps quit drinking at a higher rate, (2)
i.e., $box_{T\%} - box_{C\%} > 0\%$ (1)

2,

T

sample size T = 183
sample T% = 27/183 = 14.75% (2)
$SD_T = \sqrt{0.1475 \cdot 0.8525} = 0.355$ (2)
$SE_{sum_T} = \sqrt{183} \cdot 0.355 = 4.80$ (2)
$SE_{\%T} = \frac{4.80}{183} \cdot 100\% = 2.62\%$ (2)

C

sample size C = 188
sample C% = 6/188 = 3.19% (2)
$SD_C = \sqrt{0.0319 \cdot 0.9681} = 0.176$ (2)
$SE_{sum_C} = \sqrt{188} \cdot 0.176 = 2.41$ (2)
$SE_{\%C} = \frac{2.41}{188} \cdot 100\% = 1.28\%$ (2)

$SE_{diff\%} = \sqrt{(2.62\%)^2 + (1.28\%)^2} = 2.92\%$ (3)

$z = \frac{14.75\% - 3.19\%}{2.92\%} = 3.96$ (3)

3, area between −3.95 and 3.95: 99.992% (3)
P-value = area above 3.95
$= \frac{100\% - 99.992\%}{2} = 0.004\%$ (3)

4, • reject the null (2)
• result is highly stat. significant (2)
• T helps quit drinking at a higher rate (2)