Unit 7 Practice Test

Name Solutions

1. Suppose you were asked to analyze each of the situations described below. (NOTE: DO NOT DO THESE PROBLEMS!) For each, indicate which inference procedure you would use (from the list), the test statistic (z, t, or χ^2), and, if t or χ^2, the number of degrees of freedom.

	Type	z, t, or χ^2	df
a.	5	t	39
b.	6	χ^2	3
c.	3 (7, 8)	z, (χ^2, χ^2)	–, (1, 1)
d.	4	t	24
e.	9	t	7
f.	3	t	21
g.	7	χ^2	3
h.	1	z	—

1. proportion, 1 sample
2. difference of proportions, 2 samples
3. mean, 1 sample
4. mean of differences, matched pairs
5. difference of means, independent samples
6. goodness of fit
7. homogeneity
8. independence
9. regression, inference for β

a. A researcher wonders if meat in the diet may be a factor in high blood pressure. She compares the blood pressures of 40 randomly selected vegetarians, to those of 40 people who eat meat. means, difference, independent samples

b. According to the American Red Cross, 45% of Americans have Type O blood, 40% Type A, 11% Type B, and 4% Type AB. Last week a blood drive at the high school collected 132 pints of blood. If 51 were Type O, 55 Type A, 17 Type B, and 9 were Type AB, was this yield unusual in any way? χ^2-GOF $df = 4-1 = 3$

c. Among a random sample of college-age drivers 5% of the 576 men said they had been ticketed for speeding during the past year, compared to only 3% of the 552 women. Does this indicate a significant difference between college males and females in terms of being ticketed for speeding? difference of proportions, indep. samples (or χ^2 homogeneity or indep)

d. Who is paid more in New York State - teachers or policemen? We select a random sample of 25 New York cities and find the starting salaries of teachers and policemen in each. ~~pairs~~ teacher/police for each

e. Researchers offer small cookies to nine nursery school children and record the number of cookies consumed by each. Forty-five minutes later they observe these children during recess, and rate each child for hyperactivity on a scale from 1 – 20. Is there any evidence that sugar contributes to hyperactivity in children? relation between values, regression for regression $df = n-2 = 7$

f. 22 people complaining of indigestion take an antacid. They report that their discomfort subsided in an average of 13 minutes; the standard deviation was 4 minutes. The manufacturer wants a 95% confidence interval for the "relief time". mean, 1 sample

g. A sports fan selected a random sample of 100 games from each of the NBA, the NFL, the NHL, and Major League Baseball to see if overtimes (or extra innings) are equally likely to occur in all four sports. 4 populations 1 variable comparing % χ^2-homogeneity $df = 4-1 = 3$

h. A teacher believes that no more than 10% of high school students ever cheat on an exam, but a confidential survey found that 14 of 88 randomly selected students admitted having cheated at least once. Is this strong evidence that the teacher was wrong? proportion, 1 sample

2. **Cloning** A random sample of 800 adults was asked the following question: "*Do you think current laws concerning the use of cloning for medical research are too strict, too lenient, or about right?*" The pollsters also classified the respondents with respect to highest education level attained: high school, 2-year college degree, 4-year degree, or advanced degree. We wish to know if attitudes on cloning are related to education level. (All the conditions are satisfied – don't worry about checking them.)

	Strict	Lenient	Right	Total
High school	93 106.01	107 87.38	182 188.61	382
2-year	27 28.31	19 23.33	56 50.36	102
4-year	82 75.48	50 62.22	140 134.30	272
Adv. degree	20 12.21	7 10.07	17 21.73	44
Total	222	183	395	800

$\chi^2 =$ 1.60 + 4.40 + 0.23 +
0.06 + 0.80 + 0.63 +
0.56 + 2.40 + 0.24 +
4.97 + 0.93 + 1.03 = 17.86

P = 0.0066

a. Write appropriate hypotheses.

One population, 2 variables
χ^2 – test of independence

H_0: Opinions on cloning are independent of education level.

H_A: Opinions on cloning are not independent of education level (there is an association).

b. Suppose the expected counts had not been given. Show how to calculate the expected count in the first cell (106.01).

(expected ?) · (column total)
$\frac{382}{800} \cdot 222 = 106.005$

c. How many degrees of freedom?

$df = (\#rows - 1) \cdot (\#cols - 1) = (4-1)(3-1) = (3)(2) = 6$

d. State your complete conclusion in context.

With $\alpha = .05$, p-value of .0066 is low so we reject H_0. This is sufficient statistical evidence to conclude that opinions on cloning laws are not independent of education level.

3. **Car reliability** A consumer group assigned 62 car models reliability ratings of 1 – 5 based upon repair records. They wondered if more expensive cars might be more reliable. To find out, they created the regression analysis shown. (SHOW WORK. Don't bother writing hypotheses, and you may assume the assumptions for inference were all satisfied.)

Dependent variable is:	Reliability	
Variable	Coefficient	s.e. of coeff
Constant	2.7029	0.3508
Price	0.5099	0.4116

a. df = 60 $(n-2)$

$t = 1.24$

P = .110

b. State your conclusion.

$t = \frac{b}{s_b} = \frac{0.5099}{0.4116}$

$t = 1.239$

p-value = tcdf(1.239, 999, 60)
= .110

With $\alpha = .05$, p-value = .110 is high so we fail to reject H_0. We do not have evidence that expensive cars are more reliable.

4. **College admissions** According to information from a college admissions office, 62% of the students there attended public high schools, 26% attended private high schools, 2% were home schooled, and the remaining students attended schools in other countries. Among this college's Honors Graduates last year there were 47 who came from public schools, 29 from private schools, 4 who had been home schooled, and 4 students from abroad. Is there any evidence that one type of high school might better equip students to attain high academic honors at this college? Test an appropriate hypothesis and state your conclusion.

H_0: Distribution of school type for graduates is same as for the whole college

H_a: Distribution of school type for graduates is different than the whole college.

Conditions: counts? yes ✓ SRS? assume representative expected counts ≥ 5? No, home schooled too low,

	public	private	home	abroad	
observed:	47	29	4	4	= 84
expected	52.08	21.84	1.68	8.4	
	.62(84)	.26(84)	.02(84)	.10(84)	

But we could group home + abroad together:

	public	private	"other"	
observed:	47	29	8	now, all conditions are met.
expected:	52.08	21.84	10.08	
	.62(84)	.26(84)	.12(84)	

χ^2 – goodness of fit test $df = 2$

$\chi^2 = 3.272$ p-value = .195

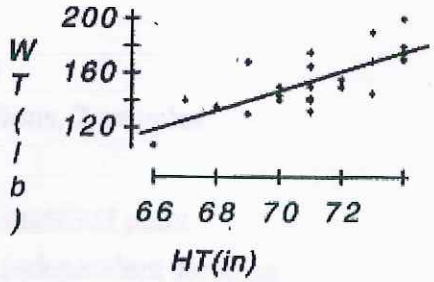
with $\alpha = .05$, p-value = .195 is high so we fail to reject H_0.

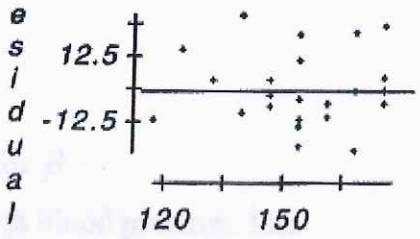
We do not have evidence that graduates from the honors program came from different types of high schools compared to the whole college.

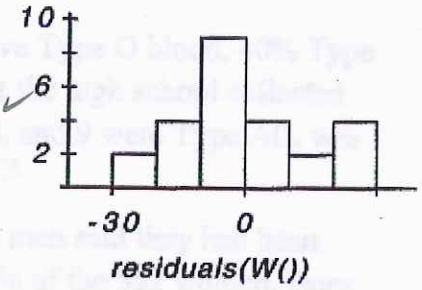
5. **Height and weight** Last fall, as our first example of correlation, we looked at the heights and weights of some AP Statistics students. Here are the scatterplot, the residuals plot, a histogram of the residuals, and the regression analysis for the data we collected from the males. Use this information to analyze the association between heights and weights of teenage boys.

Dependent variable is: WT(lb)
R squared = 56.6%
s = 14.16 with 25 - 2 = 23

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Const	-364.403	94.61	-3.85	0.0008
HT(in)	7.29993	1.333	5.48	≤ 0.0001







a. Is there an association? Write appropriate hypotheses.

H_0: Height and weight of boys are independent ($\beta = 0$)
H_A: There is an association between height and weight of boys. ($\beta \neq 0$)

b. Are the assumptions for regression satisfied? Explain.

Straight enough? scatterplot looks linear ✓
independent? residuals show no pattern ✓
variance consistent? no fanning in scatterplot or residuals ✓
residuals nearly Normal? resid. histogram good ✓

c. What do you conclude?

$t = 5.48$ for slope and p-value $\leq .0001$
with $\alpha = .05$, p-value of .0001 is low so we reject H_0.
There is sufficient statistical evidence of an association between weight and height.

d. Create a 95% confidence interval for the true slope. t^* for 95%, df = 23 = 2.069

$$CI = b \pm t^* SE_b$$
$$= 7.29993 \pm 2.069(1.333)$$
$$= (4.54, 10.06)$$

e. Explain in context what your interval means.

We are 95% confident the slope of the LSRL relating weight to height is between 4.54 and 10.06 lbs per inch.

—or—

We are 95% confident that teenage boys gain an average of between 4.54 and 10.06 pounds for every additional inch in height.