Extra Practice Problems Taken From Triola, *Elementary Statistics*, 10th Edition

Identifying H_0 and H_a

1. More than 25% of Internet users pay bills online
2. Most households have telephones
3. The mean weight of women who won Miss America titles is equal to 121 lb.
4. The percentage of workers who got a job through their college is no more than 2%.
5. Plain M&M candies have a mean weight that is at least 0.8535 g.
6. The success rate with surgery is better than the success rate with splinting.
7. Unsuccessful job applicants are from a population with a greater mean age than the mean age of successful applicants.

Example: Random Generation of Data
The TI-83/84 Plus calculator can be used to generate random data from a normally distributed population. The command **randNorm(100,15,50)** generates 50 values from a normally distributed population with $\mu = 100$ and $\sigma = 15$. One such generated sample of 50 values has a mean of 98.4 and a SD of 16.3. Use a 0.10 significance level to test the claim that the sample actually does come from a population with a mean equal to 100. Assume that σ is known to be 15. Based on the results, does it appear that the calculator's random number generator is working correctly?

Example: Umpire Strike Rate
In a recent year, some professional baseball players complained that umpires were calling more strikes than the average rate of 61.0% called the previous year. At one point in the season, umpire Dan Morrison called strikes in 2231 of 3581 pitches (based on data from *USA Today*).

a) Using $\alpha = 0.05$, test the claim that his strike rate is greater than 61.0%.

b) Construct a 95% CI for the proportion of strikes called by umpire Dan Morrison. Interpret the interval. How do the results compare with the test in part (a)?

Example: Adverse Effect of Clarinex
The drug Clarinex is used to treat symptoms from allergies. In a clinical trial of this drug, 2.1% of the 1655 treated subjects experienced fatigue. Among the 1652 subjects given placebos, 1.2% experienced fatigue. Use a 0.05 significance level to test the claim that the incidence of fatigue is greater among those who use Clarinex. Does fatigue appear to be a major concern for those who use Clarinex?

Example: Racial Profiling
Racial profiling is the controversial practice of targeting someone for suspicion of criminal behavior on the basis of race, national origin, or ethnicity. The table below includes data from randomly selected drivers stopped by police in a recent year (based on data from the U.S. Department of Justice, Bureau of Justice Statistics).

a) Use a 0.05 significance level to test the claim that the proportion of blacks stopped by police is significantly greater than the proportion of whites.

b) Construct a confidence interval that could be used to test the claim in part (a). Be sure to use the correct level of significance. What do you conclude based on the confidence interval?

	Race and Ethnicity	
	Black and Non-Hispanic	**White and Non-Hispanic**
Drivers stopped by police	24	147
Total number of observed drivers	200	1400

Example: Readability of J. K. Rowling and Leo Tolstoy

Listed below are Flesch Reading Ease scores taken from randomly selected pages in J. K. Rowling's *Harry Potter and the Sorcerer's Stone* and Leo Tolstoy's *War and Peace*.

Higher Flesch Reading Ease scores indicated writing that is easier to read. Use a 0.05 significance level to test the claim that *Harry Potter and the Sorcerer's Stone* is easier to read than *War and Peace*. Is the result as expected?

Rowling	85.3	84.3	79.5	82.5	80.2	84.6	79.2	70.9	78.6	86.2	74.0	83.7
Tolstoy	69.4	64.2	71.4	71.6	68.5	51.9	72.2	74.4	52.8	58.4	65.4	73.6

Example: Self-Reported and Measured Heights of Male Statistics Students
Eleven male statistics students were given a survey that included a question asking them to report their height in inches. They weren't told that their height would be measured, but heights were accurately measured after the survey was completed. Anonymity was maintained through the use of code numbers instead of names, and the results are shown below. Is there sufficient evidence to support a claim that male statistics students exaggerate their heights?

Reported Height	68	74	66.5	69	68	71	70	70	67	68	70
Measured Height	66.8	73.9	66.1	67.2	67.9	69.4	69.9	68.6	67.9	67.6	68.8

Example: Confidence Interval and Hypothesis Test for Bipolar Depression Treatment
In clinical experiments involving different groups of independent samples, it is important that the groups be similar in the important ways that affect the experiment. In an experiment designed to test the effectiveness of paroxetine for treating bipolar depression, subjects were measured using the Hamilton depression scale with the results given below (based on data from "Double-Blind, Placebo-Controlled Comparison of Imipramine and Paroxetine in the Treatment of Bipolar Depression," by Nemeroff et al., *American Journal of Psychiatry*, Vol. 158, No. 6).

b)   Construct a 95% CI for the difference between the two population means. Based on the results, does it appear that the two populations have different means? Should paroxetine be recommended as a treatment for bipolar depression?

c)   Use a 0.05 significance level to test the claim that the treatment group and placebo group come from populations with the same mean. How does this result compare with the CI found in part (a)?

	n	$\bar{x}$	s
Placebo	43	21.57	3.87
Treatment	33	20.38	3.91

Good Problems from Moore, McCabe and Craig, 8th Edition

Chapter 9 Exercises starting on page 557:
9.40 (d) only

 9.48

9.50

Chapter 10 Exercises starting on page 600

A regression problem. The premise for this problem is 10.32 from you book. I don't want you to do the problem as written; rather, I want you to be able to do parts (a) – (f) listed below. I have provided some partial R output for the regression of IBI (y) on Area (x) [this is what could show up on the final]. Make sure you know how to read the output to answer the specified parts of the problem. I have highlighted numbers you need in yellow. NOTE: Residual standard error of 16.53 is s_e; 47 degrees of freedom is equal to $n - 2$; Multiple R-squared of 0.1988 is the coefficient of determination.

```
Call:
lm(formula = IBI ~ Area, data = ibi.dat)

Coefficients:
            Estimate  Std. Error
(Intercept)  52.9230     4.4835
Area          0.4602     0.1347

Residual standard error: 16.53 on 47 degrees of freedom
Multiple R-squared:  0.1988
```

a) Give the statistical model for simple linear regression for this problem.

b) What is the regression equation for IBI on Area?

c) Interpret the slope in the context of the problem.

d) State the null and alternative hypotheses for examining the relationship between IBI and area.

e) Conduct the test for the hypothesis specified in (d). Interpret your result in the context of the problem.

f) Interpret the coefficient of determination in the context of the problem.

10.36—use the regression output from above and any relevant calculations from above to do the calculations for this problem by hand. Note that $SS_{xx} = 15062$ and $\bar{x} = 28.29$.