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.

Statement about the alternative; $H_0: p = 0.25\ versus\ H_a: p > 0.25$

2. Most households have telephones

Statement about the alternative; $H_0: p = 0.50\ versus\ H_a: p > 0.50$

3. The mean weight of women who won Miss America titles is equal to 121 lb.

Statement about the null; $H_0: \mu = 121\ versus\ H_a: \mu \neq 121$

4. The percentage of workers who got a job through their college is no more than 2%.

Statement about the null $H_0: p = 0.02\ versus\ H_a: p > 0.02$

5. Plain M&M candies have a mean weight that is at least 0.8535 g.

Statement about the null; $H_0: \mu = 0.8535\ versus\ H_a: \mu < 0.8535$

6. The success rate with surgery is better than the success rate with splinting.

Statement about the alternative; Let S = surgery and P = splinting; $H_0: p_S = p_P\ versus\ H_a: p_S > p_P$

7. Unsuccessful job applicants are from a population with a greater mean age than the mean age of successful applicants.

Statement about the alternative; Let U = unsuccessful and S = successful; $H_0: \mu_U = \mu_S\ versus\ H_a: \mu_U > \mu_S$

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?

Given: $n = 50;\ \bar{x} = 98.4;\ s = 16.3;\ \sigma = 15;\ \alpha = 0.10.$

$H_0: \mu = 100\ versus\ H_a: \mu \neq 100$

$z_{crit} = z_{\alpha/2} = z_{0.05} = \pm 1.645$

$$z = \frac{98.4 - 100}{\frac{15}{\sqrt{50}}} = \frac{-1.6}{2.121} = -0.75$$

p-value = $2(P(Z < -0.75)) = 2(0.2266) = 0.4532$

Since $p > \alpha$ fail to reject the null. There is insufficient evidence to suggest that the mean value from the random number generator is different than 100. There is no evidence of problems with the calculator's random number generator.

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%.

Given: $n = 3581$; $\hat{p} = \frac{2231}{3581} = 0.623$; $\alpha = 0.05$.

$H_0: p = 0.61 \text{ versus } H_a: p > 0.61$

$z_{crit} = z_\alpha = z_{0.05} = 1.645$

$$z = \frac{0.623 - 0.61}{\sqrt{\frac{0.61(0.39)}{3581}}} = \frac{0.013}{0.0082} = 1.59$$

p-value $= P(Z > 1.59) = 1 - P(Z < 1.59) = 1 - 0.9441 = 0.0559$.

Since $p > \alpha$ fail to reject the null. There is insufficient evidence to indicate that the proportion of strikes called by umpire Dan Morrison exceeds the average rate of 61%.

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)?

$$SE(\hat{p}) = \sqrt{\frac{0.623(0.377)}{3581}} = 0.0081$$

$z_{crit} = z_{\alpha/2} = z_{0.025} = 1.96$

$0.623 \pm (1.96)(0.0081) \Longrightarrow 0.623 \pm 0.0159 \Longrightarrow (0.607, 0.639)$

We are 95% confident that the proportion of strikes called by umpire Dan Morrison falls between about 60.7% and 63.9%. The interval contains 61% so 61%, the null hypothesized value in part (a), could be a likely value for the proportion of strikes. The confidence interval, however, does not completely correspond to the hypothesis test in part (a) since the test was one-tailed. You need to construct a 90% CI in order for the CI to be essentially equivalent to 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?

Let T = treatment (Clarinex) and C = control (Placebo).

Given: $n_T = 1655$; $\hat{p}_T = 0.021$; $n_C = 1652$; $\hat{p}_C = 0.012$; $\alpha = 0.05$.

$H_0: p_T = p_C \text{ versus } H_a: p_T > p_C$

$z_{crit} = z_\alpha = z_{0.05} = 1.645$

$$\bar{p} = \frac{1655(0.021) + 1652(0.012)}{1655 + 1652} = 0.017$$

$$z = \frac{(0.021 - 0.012) - 0}{\sqrt{\frac{0.017(0.983)}{1655} + \frac{0.017(0.983)}{1652}}} = 2$$

p-value $= P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$.

Since $p < \alpha$ reject the null. There is sufficient evidence to indicate that the incidence of fatigue is greater for Clarinex users. However, one has to consider practical significance versus statistical significance. Although the two proportions are statistically significantly different, is fatigue a major concern? Perhaps not since the overall rate is low; the number of people that may be impacted is low relative to the population. Of course, everything really depends on the extent of the fatigue, etc.

3

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).

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

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.

Let B = blacks and W = whites.

Given: $n_B = 200;\ \hat{p}_B = 0.12;\ n_W = 1400;\ \hat{p}_W = 0.105;\ \alpha = 0.05.$

$H_0: p_B = p_W \ versus\ H_a: p_B > p_W$

$z_{crit} = z_\alpha = z_{0.05} = 1.645$

$$\bar{p} = \frac{24 + 147}{200 + 1400} = 0.107$$

$$z = \frac{(0.12 - 0.105) - 0}{\sqrt{\frac{0.107(0.893)}{200} + \frac{0.107(0.893)}{1400}}} = 0.64$$

p-value = $P(Z > 0.64) = 1 - P(Z < 0.64) = 1 - 0.7389 = 0.2611.$

Since $p > \alpha$ fail to reject the null. There is no evidence to suggest that the proportion of blacks stopped by police is 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?

Need to construct a 90% CI to test the one-tailed alternative hypothesis in part (a).

$$SE(\hat{p}_B - \hat{p}_W) = \sqrt{\frac{0.12(0.88)}{200} + \frac{0.105(0.895)}{1400}} = 0.0244$$

$z_{crit} = z_{\alpha/2} = z_{0.05} = 1.645$

$0.015 \pm (1.645)(0.0244) \Longrightarrow 0.015 \pm 0.0401 \Longrightarrow (-0.025, 0.055)$

We are 90% confident that the difference in the proportions of blacks and whites stopped by police falls between negative 2.5% and 5.5%. Since the interval contains zero, it is possible that the two proportions are equal.

4

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

You need to start by calculating summary statistics for Rowling and Tolstoy. Let R = Rowling and T = Tolstoy.
Given: $n_R = n_T = 12$; $\bar{x}_R = 80.75$; $s_R = 4.681$; $\bar{x}_T = 66.15$; $s_T = 7.858$; Assume $\alpha = 0.05$.
$H_0: \mu_R = \mu_T \ versus \ H_a: \mu_R > \mu_T$
Since the two sample standard deviations are quite different, I will use an upooled variance (note that the "Rule of Thumb" for variances indicate the ratio is less than 3, but not much less so I am choosing unpooled). I will use $\min(n_R - 1, n_T - 1) = 11$ for the degrees of freedom.
$t_{crit} = t_{\alpha,df} = t_{0.05,11} = 1.80$

$$t = \frac{(80.75 - 66.15) - 0}{\sqrt{\frac{4.681^2}{12} + \frac{7.858^2}{12}}} = 5.53$$

p-value = $P(T_{11} > 5.53) < 0.001$

Since $p < \alpha$ reject the null. There is strong evidence that the mean Flesch Reach Ease score for Rowling's book is greater than the mean score for Tolstoy's book, suggesting that *Harry Potter and the Sorcerer's Stone* is easier to read than *War and Peace*.

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

This problem was mostly done in lecture (in class and online). The part that isn't shown in the slides is calculating the summary statistics. You first have to calculate the d_i's by taking Reported Height minus Measured Height for each individual. Then calculate $\bar{d}$ as the mean of the d_i's and s_d as the sample standard deviation of the d_i's.

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).

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

a) 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?

Using the values calculated in part (b) below, the 95% CI is $1.19 \pm 2(0.8996) \Rightarrow (-0.609, 2.989)$. We are 95% confident that the mean difference in Hamilton depression scores for Paroxetine and Placebo subject is between negative 0.609 and 2.989. Since zero is contained in the interval, it is quite possible that the means of the two groups are equal.

b) 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)?

Let P = Placebo and T = Treatment.
Given: summary statistics in the table above and $\alpha = 0.05$.
$H_0: \mu_P = \mu_T \ versus \ H_a: \mu_P \neq \mu_T$
Since the two sample standard deviations are quite similar, I will use a pooled variance with $df = 43 + 33 - 2 = 74$.
$t_{crit} = t_{\alpha/2,df} = t_{0.025,74} = 2.0$

$$s_p = \sqrt{\frac{(43-1)(3.87)^2 + (33-1)(3.91)^2}{43+33-2}} = 3.887$$

$$t = \frac{(21.57 - 20.38) - 0}{3.887 \times \sqrt{\frac{1}{43} + \frac{1}{33}}} = 1.32$$

p-value = $2(P(T_{74} > 1.32)) < 0.198$

Since $p > \alpha$ fail to reject the null. There is no evidence that the mean Hamilton depression scores for subjects taking Paroxetine is different than the mean score for subjects taking a Placebo. Paroxetine should not be recommended as a treatment for bipolar depression. The failure to reject the null hypothesis is consistent with observing zero in the 95% CI calculated in part (a).

Chapter 9 Problems from Moore, McCabe and Craig

9.40d

We have

C_M = # male students in sample = 3228 + 9659 = 12887,
C_F = # female students in sample = 10295 +4620 = 14915,
C_Y = # students in sample that lied = 3228 + 10295 = 13523,
C_N = # students in sample that didn't lie = 9659 + 4620 = 14279, and
$n = C_M + C_F$ = 12887 + 14915 = 27802

and

$O_{M,Y}$ = Observed count of male students that lied = 3228,
$O_{F,Y}$ = Observed count of female students that lied = 10295,
$O_{M,N}$ = Observed count of male students that didn't lie = 9659, and
$O_{F,N}$ = Observed count of female students that didn't lie = 4620,

and

$E_{M,Y}$ = Expected count of male students that lied = $\frac{C_M C_Y}{n} = 6268.29$,

$E_{F,Y}$ = Expected count of female students that lied = $\frac{C_F C_Y}{n}$ = 7254.71,

$E_{M,N}$ = Expected count of male students that didn't lie = $\frac{C_M C_N}{n}$ = 6618.71, and

$E_{F,N}$ = Expected count of female students that didn't lie = $\frac{C_F C_Y}{n}$ = 7660.29.

We can now compute the χ^2 statistic as $\chi^2 = \sum_{i\in\{M,F\},j\in\{Y,N\}} \frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}} = 5351.94$

We now compute the p-value of this test statistic using a χ^2-distribution with $(2-1)(2-1) = 1$ degree of freedom. As our χ^2 is very large my software simply reports the p-value equaling 0 (so the p-value is so small it is beyond the machine's ability to accurately represent). This provides very strong evidence that gender and whether a student lies are related, we would certainly reject the null hypothesis and conclude that gender and a student's lying status are statistically dependent.

9.48

(9.37) The problem has is looking to compare the distributions of "claims allowed" for three different claim sizes. Clearly this looks to compare different distributions and is thus of the first type.

(9.39) Again we have that the distributions of three different populations are being com- pared, hence this is of the first type.

(9.40) This problem would like to determine whether there is an association between gen- der and having lied to the teacher when taking one sample of students from the population. Thus it is of the second type.

(9.42) A single survey was conducted from which wish to determine if there is an association between "field of study" and "using government loads." It follows that this problem is of the second type.

(9.45) From a single survey we wish to determine if there is a relationship between bed- time and waking symptoms. Hence this is of the second type.

9.50

Let X have a Normal(0,1) distribution. Then we have that

$P(X < -0.6) = 0.2743$,
$P(-0.6 \leq X < -0.1) = P(X < -0.1) - P(X < -0.6) = 0.1859$,
$P(-0.1 \leq X < 0.1) = P(X < 0.1) - P(X < 0.6) = 0.07966$,
$P(0.1 \leq X < 0.6) = P(-0.6 \leq X < -0.1) = 0.1859$, and
$P(0.6 \leq X) = P(X < -0.6) = 0.2743$,

where for the last two equalities we used the symmetry of the normal distribution. We now compute our goodness-of-fit test statistic as

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 3.4$$

Using this test statistic, we use a χ^2-distribution with 5 − 1 = 4 degrees of freedom to compute the p-value 0.49. As this p-value is quite large we fail to reject the null hypothesis that the probability of falling into the intervals agreed with the probability of the standard normal distribution being in said intervals. There is no evidence that the data are different than that which would be observed under a normal distribution.

Regression Problems

```
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.

Let IBI_i be the ith IBI observation and $Area_i$ be the ith area observation. Then, for all $1 \leq i \leq n = 49$, we consider the simple linear regression model

$$IBI_i = \beta_0 + \beta_1 \cdot Area_i + \epsilon_i$$

where β_0 and β_1 area unknown parameters and $\epsilon_1, \ldots, \epsilon_n$ are independent samples from a normal distribution with mean 0 and (unknown) standard deviation σ.

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

This regression gives us estimates $b_0 = 52.923$, $b_1 = 0.4602$, so that our estimated regression line is $\hat{y} = 52.923 + 0.4602(Area)$

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

For each 1 km increase in area, the IBI will increase by 0.46 units, on average.

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

$H_0: \beta_1 = 0$ versus $H_a: \beta_1 \neq 0$

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

$t = \frac{0.4602}{0.1347} = 3.42$; p-value is approximately zero. There is sufficient evidence to reject the null; the slope for area is statistically different than zero. In fact, the slope appears to be greater than zero.

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

Approximately 20% of the variation in IBI can be explained by knowing area.

10.36

(a) Recall our definition of $\widehat{\mu}_{IBI}(Area)$ from the prior problem. We wish to compute a confidence interval of $\mu_{IBI}(40)$ (the mean of IBI when area is fixed at 40km^2) for which

we have point estimate $\widehat{\mu}_{IBI}(40) = 52.92 + 0.46 \cdot 40 = 71.32$. Recall that the standard error of $\widehat{\mu}_{IBI}(40)$ equals

$$SE_{\widehat{\mu}_{IBI}(40)} = s\sqrt{\frac{1}{n} + \frac{(40 - \overline{Area})^2}{\sum_{i=1}^{n}(Area_i - \overline{Area})^2}}$$

$$SE_{\widehat{\mu}_{IBI}(40)} = 16.53\sqrt{\frac{1}{49} + \frac{(40 - 28.29)^2}{15062}} = 2.84.$$

Letting $t^* = 2.01$ be the 97.5th quantile of a $t(n-2) = t(47)$ distribution we now have 95% confidence interval for the mean response when $Area = 40$ equaling

$$(71.32 - 2.01 \cdot 2.84, 71.32 + 2.01 \cdot 2.84) = (65.61,\ 77.03).$$

Hence we are 95% confident that the mean IBI for streams in the Ozark Highland ecoregion of Arkansa with an area of 40km^2 is between the values 65.61 and 77.03.

(b) The computations here are similar as for (a) but instead of using $SE_{\widehat{\mu}_{IBI}}(40)$ we use $SE_{\widehat{y}}$, the standard error for predicting an individual response $\widehat{y}$. We have that

$$\begin{aligned} SE_{\widehat{y}} &= s\sqrt{1 + \frac{1}{n} + \frac{(40 - \overline{Area})^2}{\sum_{i=1}^{n}(Area_i - \overline{Area})^2}} \\ &= 16.53\sqrt{1 + \frac{1}{49} + \frac{(40 - 28.29)^2}{15062}} \\ &= 16.77. \end{aligned}$$

Hence we have 95% prediction interval

$$(71.32 - 2.01 \cdot 16.77, 71.32 + 2.01 \cdot 16.77) = (37.61,\ 105.03),$$

note that this interval is much larger than the interval from (a). This confidence interval tells us that we can be 95% confident that if we were to randomly select a stream in the Ozark Highland ecoregion of Arkansas with an area of 40km^2 then it would have IBI between 37.61 and 105.3.

(c) See parts (a) and (b) for these interpretations.

(d) It seems very unlikely that these results would directly generalize directly to other streams in Arkansas or in other states. Presumably the effect of area on water quality in streams is highly dependent on the local ecosystem and surrounding human activity. If, for example, the Ozark Highland ecoregion has soil that traps many pollutants then presumably the effect of area on IBI would be greater in the Ozark region than in a region with poor soil.

■