True/false questions.

**Problem 1.1.** (2 points)
The higher the slope of the least-squares regression line, the stronger the association between the explanatory and the response variables.

*True or false?*

**Solution:** FALSE

**Problem 1.2.** (2 points)
If a random variable $X$ has a standard normal distribution, then $X^2$ has a chi-squared distribution with 1 degree of freedom.

*True or false?*

**Solution:** TRUE

**Problem 1.3.** (2 points)
We can use the $F$–test to test whether the population standard deviations are equal for two populations.

*True or false?*

**Solution:** TRUE

Free-response problems.

**Problem 1.4.** (10 points)
A die is rolled 60 times and the face values are recorded. The results are as follows:

<table>
<thead>
<tr>
<th>Up face</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of occurrences</td>
<td>8</td>
<td>11</td>
<td>0</td>
<td>12</td>
<td>20</td>
<td>9</td>
</tr>
</tbody>
</table>

At the 5% significance level, test the hypothesis that the die is fair.

**Solution:**
Problem 1.5. (10 points)
Source: Problem 8.99 from the textbook.

*Castaneda v. Partida* is an important court case in which statistical methods were used as part of a legal argument. When reviewing this case, the Supreme Court used the phrase two or three standard deviations as a criterion for statistical significance. This Supreme Court review has served as the basis for many subsequent applications of statistical methods in legal settings. (The two or three standard deviations referred to by the Court are values of the $z$ statistic and correspond to $p$-values of approximately 0.05 and 0.0026.)

In Castaneda the plaintiffs alleged that the method for selecting juries in a county in Texas was biased against Mexican Americans. For the period of time at issue, there were 181,535 persons eligible for jury duty, of whom 143,611 were Mexican Americans. Of the 870 people selected for jury duty, 339 were Mexican Americans.

(i) (1 point) What proportion of eligible jurors were Mexican Americans?

(ii) (2 points) Let $p$ denote the probability that a randomly selected juror is a Mexican American. Formulate the null and alternative hypotheses to be tested.

(iii) (1 point) What is the sample proportion of jurors who were Mexican American?

(iv) (4 points) Compute the $z$-statistic, and find the $p$-value.

(v) (2 points) How would you summarize your conclusions? (A finding of statistical significance in this circumstance does not constitute proof of discrimination. It can be used, however, to establish a prima facie case. The burden of proof then shifts to the defense.)

Solution:
Problem 1.6. (10 points)

It is claimed that the bags of chocolate chips available in Costco contain at least 4 pounds (64 ounces).

A random sample of 50 bag measurements resulted in the sample average of 62 and the sample standard deviation of 8. Please, test the hypothesis that the bags contain at least 64 ounces of delicious chocolate chips at the significance level of 5%.

Solution:

Let $\mu$ denote the population mean of the weight of bags of chocolate chips. The hypotheses are:

$$H_0 : \mu = 64 \quad vs. \quad H_a : \mu < 64.$$

Since the population standard deviation is not given, we should use the $t$-test. However, with the sample size of 50 we can be comfortable enough using the $z$-test. The observed value of the test-statistic is

$$z = \frac{62 - 64}{8/\sqrt{50}} = \frac{-5\sqrt{2}}{4} = -1.768.$$

On the other hand, the critical value associated with a left-sided hypothesis test with a 5% confidence level is $-1.645$. Since the observed value of the test statistic falls below the critical value, we reject the null hypothesis.

Problem 1.7. (9 points)

Consider an experiment designed to study the causal relationship between temperature the crickets’ chirps. The following is the statistical software output of a simple linear regression.

Call:

lm(formula = crickets$Chirps ~ crickets$Temperature)

Residuals:

1 2 3 4 5 6 7
6.9534 1.6816 -9.3359 -6.3534 1.5019 5.8485 -0.2961

Coefficients:

Estimate Std. Error t value Pr(>|t|)
(Intercept) -157.8165 18.1059 -8.716 0.000329 ***
crickets$Temperature 4.2544 0.2624 16.215 1.63e-05 ***

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 6.561 on 5 degrees of freedom
Multiple R-squared: 0.9813, Adjusted R-squared: 0.9776
F-statistic: 262.9 on 1 and 5 DF, p-value: 1.626e-05

(i) (3 points) What is the equation of the least-squares regression line?

(ii) (4 points) What is the value of the coefficient of determination? Is this value indicative of a strong or weak association between the two variables?
(iii) (2 points) What is the estimate of the standard deviation of the regression error?

**Problem 1.8.** (15 points)
Suppose that, in our usual notation, $\hat{p}_1 = 0.5, \hat{p}_2 = 0.2, n_1 = 20$ and $n_2 = 30$. What is the $p-$value for testing $H_0 : p_1 = p_2$ vs. $H_a : p_1 \neq p_2$.

**Solution:** Under the null hypothesis, the two proportion parameters are equal. So, the estimate for the proportion parameter of the entire population is

$$\hat{p} = \frac{\hat{p}_1 \times n_1 + \hat{p}_2 \times n_2}{n_1 + n_2} = \frac{10 + 6}{50} = 0.32.$$

Thus, the observed value of the $z-$statistic under the null hypothesis is

$$z_{obs} = \frac{0.5 - 0.2}{\sqrt{0.32 \times 0.68 \times \left( \frac{1}{20} + \frac{1}{30} \right)}} \approx 2.23$$

The associated $p-$value is

$$2(1 - \Phi(2.23)) = 0.0258.$$
Multiple-choice problems.

**Problem 1.9.** (5 points) *Source: Ramachandran-Tsokos.*

A dendritic tree is a branched formation that originates from a nerve cell. In order to study brain development, researchers want to examine the brain tissues from adult guinea pigs. At least how many cells must the researchers select (randomly) so as to be 95% sure that the sample mean is within 3.4 cells of the population mean? Assume that a previous study has shown that the cells.

(a) 28  
(b) 33  
(c) 34  
(d) 35  
(e) None of the above.

**Solution:** (c)

**Problem 1.10.** (5 points)

The hypotheses $H_0 : \mu = 10$ versus $H_a : \mu \neq 10$ are examined using a sample of size $n = 18$. The one-sample $t$-statistic has the value of $t = -2.05$. Between what two values does the $P$-value of this test fall?

(a) $0.01 < P-$value $< 0.02$,  
(b) $0.02 < P-$value $< 0.025$,  
(c) $0.025 < P-$value $< 0.05$,  
(d) $0.05 < P-$value $< 0.10$  
(e) None of the above.

**Solution:** (d)