Econometrics Problem Sets
William M. Boal

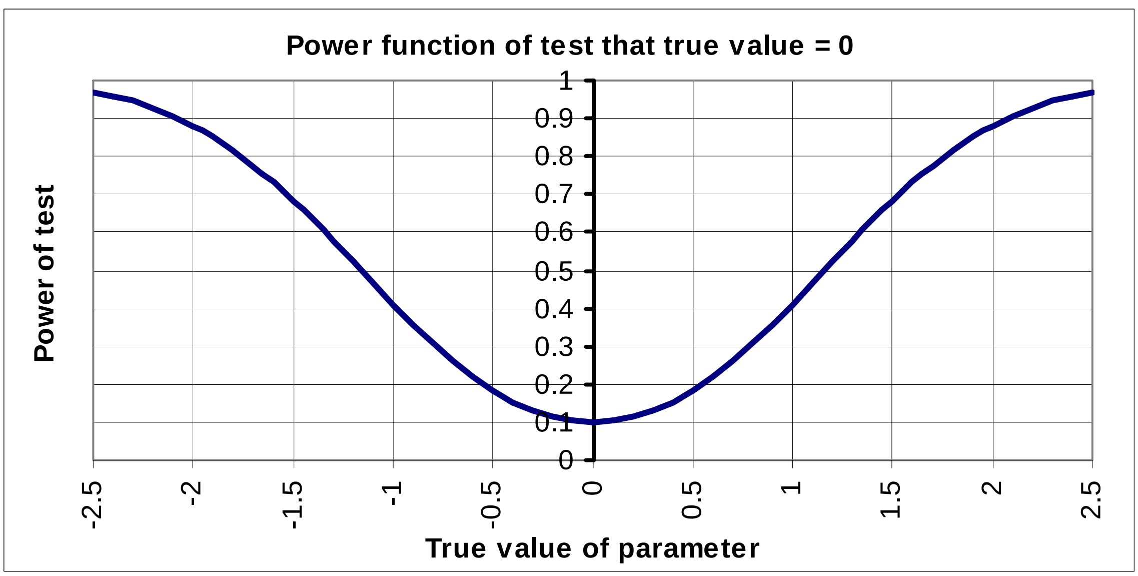
Problem Set 7
"Statistics: Confidence Intervals and Hypothesis Tests"

(7.1) [Theory of hypothesis testing] The following table shows the probabilities of various outcomes for a particular hypothesis test. Here "H_0" denotes the null hypothesis and "H_1" denotes the alternative hypothesis. Note that two probabilities are missing.

		Correct hypothesis in reality	
		H_0	H_1
Decision indicated by the test	H_0	Prob = 0.7	(b)
	H_1	(a)	Prob = 0.9

a. What probability value belongs at (a)? [Hint: Which is random according to classical statistics—reality, or the outcome of the test?]
b. What probability value belongs at (b)?
c. What value is the size or significance of the test?
d. What value is the power of the test?

(7.2) [Theory of hypothesis testing] We wish to test the null hypothesis that a certain population parameter is zero. Suppose a particular test has the power function below.



a. What is the probability that this test will reject the null hypothesis when the true value of the parameter is -2.0?
b. What is the probability that this test will reject the null hypothesis when the true value of the parameter is +1.5?
c. What is the size or significance of this test?

(7.3) [Normal population distribution, small sample] Suppose the strength of a certain material is measured $n = 15$ times. Measurement is subject to random error, so these measurements constitute a random sample, whose population mean is the true unknown strength of the material μ, and whose variance is σ^2, also unknown. Moreover it is believed that errors in measurement are normally-distributed. Thus each measurement x_i follows the normal distribution $N(\mu,\sigma^2)$. The following statistics have already been calculated. Here, $\bar{x}$ is the sample mean.

$$\sum_{i=1}^{n} x_i = 375 \quad \text{and} \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = 840$$

a. Is the normal distribution discrete or continuous? Why?
b. Compute a (best unbiased) estimate of the true or population mean μ.
c. Compute an unbiased estimate of the true or population variance σ^2.
d. Compute the standard error for your estimate of μ.
e. What are the "degrees of freedom" for this problem? (Give a number.)
f. Compute a 99% confidence interval for μ.
g. Test the null hypothesis H_0: $\mu=20$ against the one-sided alternative hypothesis H_1: $\mu>20$ at 5% significance. Give
   - the *value* of the test statistic
   - the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.

(7.4) [Normal population distribution, small sample] We seek to assess the academic achievement of a student population by administering a test to a small number of students selected at random. Suppose the test is given to **n=10** students. We assume their test scores are a random sample from a normal distribution with unknown population mean μ and unknown population variance σ^2, that is $N(\mu,\sigma^2)$. Each score (observation) is denoted x_i for i=1 through n. The following statistics have already been computed:

$$\sum_{i=1}^{n} x_i = 540 \quad \text{and} \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = 810$$

We seek estimates of the population mean and variance on the basis of these data.

a. Compute a (best unbiased) estimate of μ.
b. Compute an unbiased estimate of σ^2.
c. Compute the standard error for your estimate of μ.
d. What are the "degrees of freedom" for this problem? (Give a number.)
e. Compute a 99% confidence interval for μ.
f. Test the null hypothesis H_0: $\mu=50$ against the one-sided alternative hypothesis H_1: $\mu>50$ at 5% significance. Give
   - the *value* of the test statistic
   - the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.

(7.5) [Arbitrary population distribution, large sample] Suppose we wish to analyze the distribution of the number of children per family in a population. Let μ denote the unknown true population mean number of children per family. Observations X_i have been collected on 400 families, with the following summary values. Here, $\overline{X}$ is the sample mean.

$$\sum_{i=1}^{400} X_i = 832 \qquad\qquad \sum_{i=1}^{400} \left(X_i - \overline{X}\right)^2 = 324$$

a. Is the population distribution discrete or continuous? Justify your answer.
b. Compute an unbiased estimate of μ.
c. Compute an estimate of σ^2.
d. Compute the standard error for your estimate of μ.
e. Compute a 95% asymptotic confidence interval for μ.
f. Test the null hypothesis that $\mu = 2$ against the one-sided alternative hypothesis that $\mu > 2$, at 5% significance using an asymptotic test. Give
   - the *value* of the test statistic
   - the *critical point* from the appropriate table at the back of your textbook (or compute the *p-value* using a spreadsheet program)
   - your conclusion: whether you reject the null hypothesis at 5% significance.

(7.6) [Arbitrary population distribution, large sample] Suppose we are conducting a survey to find out the proportion of households that have high-speed internet service. Since responses will be either "yes" or "no," each random observation is a Bernoulli random variable with unknown parameter p. Let n denote the number of households in our survey, a number yet to be determined.

a. Prove that the maximum variance any Bernoulli random variable could have is 0.25. [Hint: Maximize the formula for the variance with respect to p.]
b. Show that the maximum length of a 95% confidence interval for p is $\pm\left(0.98/\sqrt{n}\right)$.
c. How many households (n) must be surveyed to ensure a 95% confidence interval that is no more than ±1% (±0.01) ?
d. How many households (n) must be surveyed to ensure a 95% confidence interval that is no more than ±2% (±0.02) ?
e. How many households (n) must be surveyed to ensure a 95% confidence interval that is no more than ±3% (±0.03) ?

(7.7) [Arbitrary population distribution, large sample] A newspaper story reports that, according to a recent poll, about 43% (0.43) of the population approves of the president's performance, with a "margin of error" of plus-or-minus 3 percentage points (±0.03). Assume that "margin of error" means a 95 percent asymptotic confidence interval. Since responses are either “yes” or “no,” each random observation is a Bernoulli random variable with unknown parameter p.

a. Use the confidence interval to calculate the standard error of the 0.43 estimate.
b. Use the 43% estimate of p to compute estimates of the population variance and standard deviation. [Hint: This is a Bernoulli random variable.]
c. Calculate the number of people included in the poll ("n"). [Hint: Substitute your answers to (a) and (b) into the definition: SE = estimated SD / $\sqrt{n}$.]
d. Calculate the number of people included in the poll that approved of the president's performance.

[end of problem set]