Lab 11: Tests of significance: standard deviations and power
STT 421: Summer, 2004
Vince Melfi

In this lab we'll investigate how a test's ability to return the correct decision is affected by the sample size, by how much difference there is between the null hypothesis and the true state of the population, and by the particular decision rule we use.

11.1 Computing the test statistic and p-value

We'll use SAS to perform the calculations for Exercise 6.39 of the text. The exercise gives "Degree of Reading Power" scores for 44 third graders in a school district. We're told to assume that the scores are a random sample from the third graders in a suburban school district, and that the population distribution of scores is approximately normal with mean μ and standard deviation $\sigma = 11$. We're asked to test a researcher's belief that the mean score of third graders in this district is higher than the national mean score, which is 32.

In this case the null hypothesis would be $H_0\colon \mu = 32$ and the alternative hypothesis would be $H_a\colon \mu > 32$.

The data are in the file `u:\msu\course\stt\421\summer04\drp.dat`, which contains both the scores and the observation number. Read in the data using the following SAS code.

```
data drp;
  infile 'u:\msu\course\stt\421\summer04\drp.dat';
  input obsnum score;
```

1. Use **proc means** to compute the mean of the scores, and then compute the value of the test statistic by hand.

   Answer: z = ____________.

2. Use the standard normal table in your book to compute the p-value of the test statistic. (Be careful to remember that this is a one-sided alternative hypothesis.)

   Answer: p-value = ____________.

Here's a SAS program that answers the above questions for you. Make sure your answers above agree with its answers, and make sure you understand the program.

```
proc univariate data = drp noprint;
  output out = drpmean mean = Mean;
  var score;

data drptest;
  set drpmean;
  nullmu = 32;
  z = (mean - nullmu)/(11/sqrt(44));
  pvalue = 1 - cdf('normal', z, 0, 1);
```

```
proc print data = drptest;

run;
```

11.2 Power, cutoff, and alternative values

Often a decision to either stick with or else reject the null hypothesis must be made. Decision rules are of the form "Reject the null hypothesis if the p-value is less than α" where α is a small number chosen in advance of seeing the data. The smaller the value of α we choose, the more protection we are seeking from false rejection of the null hypothesis. For example, $\alpha = 0.01$ provides more protection against false rejection of the null hypothesis than $\alpha = 0.05$.

But this protection comes at a cost: The smaller we make α, the less likely we are to reject the null hypothesis when it is false. In this section we investigate this phenomenon, in the process seeing how the distance of the true mean from the null mean affects our chance of correctly rejecting the null hypothesis.

We'll do all this in the context of the dataset `u:\msu\course\stt\421\summer04\power1.dat`. This dataset contains four populations: `pop10, pop11, pop12,` and `pop15`, all of which are (approximately) normally distributed with means 10, 11, 12, and 15 respectively, and standard deviation $\sigma = 5$. The following program reads in the data and computes the population means and standard deviations.

```
data power1;
  infile 'u:\msu\course\stt\421\summer04\power1.dat';
  input pop10 pop11 pop12 pop15;

proc means data = power1;

run;
```

Now we'll take 1000 independent random samples of size $n = 10$ from the dataset. Note that since we don't have an **id** statement, **proc surveyselect** will take samples from all four populations.

```
proc surveyselect data = power1 n = 10 rep=1000 out=powertest;
```

Now that the dataset `powertest` containing all the samples exists, you can leave the above line out of your programs. (This is a good idea, since the selection of the samples takes some time.)

We'll first look at the samples from `pop10` which has a population mean of 10. Our plan is to test $H_0\colon \mu = 10$ versus $H_a\colon \mu > 10$, so `pop10` satisfies the null hypothesis, i.e., we should not reject H_0 for data from this population.

1. If we use the rule "Reject H_0 if the p-value is less than 0.05," what proportion of the samples do you expect will lead to rejection of H_0?

2. If we use the rule "Reject H_0 if the p-value is less than 0.01," what proportion of the samples do you expect will lead to rejection of H_0?

The following SAS code first computes the mean of each of the 1000 samples from `pop10` and stores them in `test10means`. Then it computes z statistics for each of the samples and stores them in `zstat10`. Then, for each of the samples, it decides whether to reject H_0 or not based on the two decision rules above corresponding to $\alpha = 0.05$ and $\alpha = 0.01$, and stores the results in `decide10`. (The variable `reject05` is set to 1 if we reject using $\alpha = 0.05$ and to 0 otherwise. The variable `reject01` is set to 1 if we reject using $\alpha = 0.01$ and to 0 otherwise.) And then the program applies **proc freq** to `reject10` to compute the percentage of samples which lead to rejection of H_0 under the two decision rules.

```
proc surveyselect data = power1 n = 10 rep=1000 out=powertest;

proc univariate data = powertest noprint;
  output out = test10means mean = Mean;
  var pop10;
  by replicate;

data zstat10;
  set test10means;
  nullmu = 10;
  z = (mean - nullmu)/(5/sqrt(10));
  pvalue = 1 - cdf('normal', z, 0, 1);

data decide10;
  set zstat10;
  if (pvalue < 0.05) then reject05 = 1;
  else reject05 = 0;
  if (pvalue < 0.01) then reject01 = 1;
  else reject01 = 0;
  drop nullmu z pvalue replicate mean;

proc freq data = decide10;

run;
```

1. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.05?" Is the answer close to your guess above?

2. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.01?" Is the answer close to your guess above?

11.2.1 Population mean equal to 11

Now we'll repeat the process but using the samples from `pop11`, which has a mean of 11.

1. If we use the rule "Reject H_0 if the p-value is less than 0.05," what proportion of the samples do you expect will lead to rejection of H_0?

2. If we use the rule "Reject H_0 if the p-value is less than 0.01," what proportion of the samples do you expect will lead to rejection of H_0?

Here's the relevant program. Note that it's similar to the above program, but with "10" replaced by "11" in most places, *but not in the specification of the null hypothesis mean* `nullmu = 10` *when computing the test statistic.*

```
proc univariate data = powertest noprint;
  output out = test11means mean = Mean;
  var pop11;
  by replicate;

data zstat11;
  set test11means;
  nullmu = 10;
  z = (mean - nullmu)/(5/sqrt(10));
  pvalue = 1 - cdf('normal', z, 0, 1);

data decide11;
  set zstat11;
  if (pvalue < 0.05) then reject05 = 1;
  else reject05 = 0;
  if (pvalue < 0.01) then reject01 = 1;
  else reject01 = 0;
  drop nullmu z pvalue replicate mean;

proc freq data = decide11;

run;
```

1. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.05?" Is the answer close to your guess above?

2. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.01?" Is the answer close to your guess above?

11.2.2 Population mean equal to 12

Now repeat the above process using the samples from `pop12` which is the population with mean 12.

1. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.05?"

2. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.01?"

11.2.3 Population mean equal to 15

Now repeat the above process using the samples from `pop15` which is the population with mean 15.

1. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.05?"

2. What proportion of the samples led to rejection of H_0 using the decision rule "Reject H_0 if the p-value is less than 0.01?"

3. What general conclusion do you draw about how the chance of rejecting H_0 is related to the difference between the population mean and the null hypothesis mean of 10? Does this make sense intuitively?

4. What general conclusion do you draw about how the chance of rejecting H_0 is related to the value of α in the decision rule? Does this make sense intuitively?

11.3 The role of the sample size

The above simulations were performed with a sample size of $n = 10$. In this section we'll repeat two of them with a larger sample size of $n = 50$ and see how this affects our answers. We're still testing the same hypotheses $H_0\colon \mu = 10$ versus $H_a\colon \mu > 10$.

1. Consider the population `pop10` with mean equal to 10. Do you expect your answers to the questions about "what proportion of samples led to rejection of H_0" to change? If so, how?

2. Consider the population `pop12` with mean equal to 12. Do you expect your answers to the questions about "what proportion of samples led to rejection of H_0" to change? If so, how?

We don't need to change much to do the simulation. First we must retake the samples with $n = 50$ instead of $n = 10$. Note that we use the same name `powertest` as before for the resulting data.

```
proc surveyselect data = power1 n = 50 rep=1000 out=powertest;
```

Now rerun your programs for `pop10` and `pop12` in the new setting. The only necessary change is replacing `sqrt(10)` by `sqrt(50)` in the line that computes the z statistic.

1. Consider the population `pop10` with mean equal to 10. Did your answers to the questions about "what proportion of samples led to rejection of H_0" change much? If so, how?

2. Consider the population `pop12` with mean equal to 12. Did your answers to the questions about "what proportion of samples led to rejection of H_0" change much? If so, how?

Hopefully you saw that the chance of rejecting H_0 didn't change for `pop10`, but that the chance of rejecting H_0 increased for `pop12`. Can you explain, in non-technical terms, why this makes sense?