Tests on the population mean (known variance)

1. Is a soda-dispensing machine performing according to specification? Pepsi's dispensing machine is designed to fill bottles with exactly 2 liters of their product. To test if the machine is performing according to specification, we collect a sample of 100 "2-liter" bottles. The average quantity contained in the sample bottles is $\bar{x} = 1.985$ liters. The (population) standard deviation of the fill is known to be 0.05. Test whether the machine is in control, at the 5% level of significance.

   (a) What are the null and alternative hypotheses?

   > **Solution:** The population we are interested in is all bottles filled by the machine (past and future); the mean fill level of the population is μ, and the standard deviation of the fill level of the population is $\sigma = 0.05$. Our null and alternative hypotheses are as follows:
   >
   > $$H_0 : \mu = 2$$
   > $$H_a : \mu \neq 2.$$
   >
   > The nominal (null) value of the population mean is $\mu_0 = 2$.

   (b) What is the test statistic?

   > **Solution:** Our test statistic is based on the sample of $n = 100$ bottles. Let $\bar{x}$ denote the mean fill level from this sample. Our test statistic is
   >
   > $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 2}{(.05)/\sqrt{100}} = \frac{\bar{x} - 2}{(.005)}.$$

   (c) What is the rejection region?

   > **Solution:** Since this is a 2-sided alternative, we reject for large values of $|z|$, specifically when $|z| > z_{\alpha/2}$. Since we are testing at the 5% significance level, we have $\alpha = .05$ and $z_{\alpha/2} = 1.96 \approx 2$. We reject H_0 when
   >
   > $$|z| > 2.$$

   (d) What assumptions are you making?

   > **Solution:** We are assuming that the sample bottles were drawn independently without bias from the population.

(e) What is the result of the test?

> **Solution:** The observed value of the test statistic is
> $$\frac{1.985 - 2}{.005} = -3.$$
> Since this value is in the rejection region, we reject H_0 at significance level 5%.

2. Before Facebook's recent redesign, the mean number of ad clicks per day was 100K, and the standard deviation was 35K. In the 49 days after the redesign, the mean number of ad clicks per day was 105K. Is there significant evidence that the redesign affected ad clicks? Perform a test at the 5% level, assuming the redesign did not affect the population standard deviation.

(a) What are the null and alternative hypotheses?

> **Solution:** There are two populations of interest here. The first population is the daily ad clicks (past and future), in a (hypothetical) world where Facebook does not redesign their cite. The second population is the daily ad clicks (past and future) in the world where Facebook redesigns their cite. Let μ_0 be the mean clicks per day in the first population, and μ be the mean clicks per day. Assume that the standard deviation of the number of clicks per day is the same for both populations, σ.
>
> We use "thousands of clicks" as the units for all relevant quantities. The problem tells us that $\mu_0 = 100$ and $\sigma = 35$.
>
> The null hypothesis is that there is no difference between the two populations (i.e., the redesign had no effect); the alternative hypothesis is that the two populations have different means:
>
> $$H_0 : \mu = 100$$
> $$H_a : \mu \neq 100$$

(b) What is the test statistic?

> **Solution:** The test statistic is based on the sample, the observed clicks in the $n = 49$ days after the redesign. Let $\bar{x}$ denote the mean clicks per day in the sample. Our test statistic is
> $$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 100}{35/\sqrt{49}} = \frac{\bar{x} - 100}{5}.$$

(c) What is the rejection region?

> **Solution:** Since this is a 2-sided test, we reject when $|z|$ is large, specifically when $|z| > z_{\alpha/2}$. The problem asks for a test at level 5%, so $\alpha = .05$ and $z_{\alpha/2} = 1.96 \approx 2$. We will reject H_0 if $|z| > z_{\alpha/2}$, i.e.
> $$|z| > 2.$$

(d) What assumptions are you making?

> **Solution:** We need to assume that the sample elements were drawn independently from the population. (The population we refer to is the world in which Facebook does the website redesign, not the hypothetical world where they don't.)
>
> Admittedly, the assumption does *not* hold, since there is strong selection bias in the sample: we are sampling days right after the website redesign with higher probability

than days far into the future. For example, we have no chance of sampling a day three years into the future.

Since the assumption does not hold, we are in a bit of an awkward position. The hypothesis test may not be valid. One way in which it may not be valid is the following: it usually takes people a few weeks to adjust to a website redesign, so the ad click behavior in our sample of 49 days may not be representative of all future ad click behavior.

(e) What is the result of the test?

Solution: Despite the caveats mentioned in part (d), we will proceed with the test. Using the observed value $\bar{x} = 105$, we compute

$$z = \frac{105 - 100}{5} = 1.$$

Since $|z| \leq 2$, the statistic is not in the rejection region, so we do not reject H_0 at significance level 5%.

Tests on the population mean (unknown variance)

3. This is a variation of problem 1. A certain population has unknown mean μ and unknown standard deviation σ. You want test (at significance level 5%) the null hypothesis $H_0 : \mu = 2$ against the alternative $H_a : \mu \neq 2$. To this end, you collect a sample of size $n = 100$. The sample mean is $\bar{x} = 1.985$ and sample standard deviation is $s = 0.5$.

(a) What is the test statistic?

> **Solution:**
> $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{\bar{x} - 2}{(.05)/\sqrt{100}} = \frac{\bar{x} - 2}{(.005)}.$$

(b) What is the rejection region?

> **Solution:** Since this is a 2-sided alternative, we reject for large values of $|t|$, specifically when $|t| > t_{\alpha/2}$. We have $n - 1 = 99$ degrees of freedom. Since we are testing at the 5% significance level, we have $\alpha = .05$ and $t_{\alpha/2} \approx 2$. We reject H_0 when
>
> $$|t| > 2.$$

(c) What assumptions are you making?

> **Solution:** We are assuming that the samples are drawn independently from the population.

(d) What is the result of the test?

> **Solution:** The observed value of the test statistic is
>
> $$\frac{1.985 - 2}{.005} = -3.$$
>
> Since this value is in the rejection region, we reject H_0 at significance level 5%.

(e) How is this problem different from problem **??**?

> **Solution:** The population standard deviation is unknown, so we use a t statistic instead of a z statistic.

One-sided alternatives

4. The average nicotine content of a brand of cigarettes must be less than 0.5 mg for it to qualify as a Low Nicotine brand. The manufacturer of Lucky Strikes Cigarettes claims that it is a Low Nicotine brand. To test this claim, the FDA takes a random sample of 20 cigarettes (one pack) of Lucky Strikes. They find an average nicotine content of 0.4 mg, with a sample standard deviation of 0.2 mg. Test the manufacturer's claim, at the 1% level of significance. Assume that the nicotine measurements are normally distributed.

   (a) What are the population and the sample?

   > **Solution:** The population is the nicotine contents of all Low Nicotine cigarettes manufactured by Lucky Strike. The sample is the $n = 20$ sampled cigarettes.

   (b) What are the null and alternative hypotheses?

   > **Solution:** Let μ be the average nicotine contents (in mg) of all cigarettes in the population.
   >
   > $$H_0 : \mu = 0.5$$
   > $$H_a : \mu > 0.5$$

   (c) What is the test statistic?

   > **Solution:** Since the population standard devation is unknown, we use a t statistic:
   >
   > $$T = \frac{\bar{X} - 0.2}{S/\sqrt{n}}$$

   (d) What is the rejection region?

   > **Solution:** We reject H_0 if $T > t_\alpha$ for $\alpha = 0.01$ and $n - 1 = 19$ degrees of freedom. That is, we reject H_0 if $T > 2.539$.

   (e) What assumptions are you making?

   > **Solution:** We assume that the sample is unbiased. Also, since $n < 30$, we need to assume that the population is normally distributed.

   (f) What is the result of the test?

   > **Solution:** The observed test statistic is
   > $$t = \frac{0.4 - 0.5}{0.2/\sqrt{20}} = -2.236.$$

Since this is not in the rejection region, we do not reject the null hypothesis.