Stat 8003, Homework 7

Group G: `sample( c( "David" , "Andrew", "Salam" ))`

October 14, 2014

Question 7.1. We want to know the mean percentage of butterfat in milk produced by a farm by sampling multiple loads of milk. Previous records indicate the average percent butterfat in milk is 3.35 and the standard deviation among loads is 0.15. Now we hope to detect a change of the percent butterfat in milk.

(a) Find the rejection region at the significance level $\alpha = 0.05$;

(b) Suppose 100 loads of milk are sampled. What is the power for the test for detecting a change of the mean to 3.40.

(c) Plot the power as a function of the absolute value of the change of the mean over the standard deviation (which is $|\mu_1 - \mu_0|/\sigma$).

(d) Now we hope to detect a change of the percent butterfat in milk to 3.40 with a power 0.8. How many loads do we need to sample?

Answer:

(a) Let `r.v.` X be the percentage of butterfat in milk. We have

$$X \sim N(\mu = 3.35, \sigma = 0.15)$$

Or

$$Z = \frac{X - 3.35}{0.15} \sim N(0, 1)$$

The null hypothesis is that there is no change in butterfat; the alternative hypothesis is that there is:

$$H_0 : \ \mu = 3.35$$
$$H_a : \ \mu \neq 3.35$$

The rejection region of this two sided test is given by

$$
\begin{aligned}
R &= \left\{ \, |Z| > z_{\alpha/2} \right\} \\
&= \left\{ Z < z_{0.975} \right\} \cup \left\{ Z > z_{0.025} \right\} \\
&= \left\{ \frac{X - 3.35}{0.15} < -z_{0.025} \right\} \cup \left\{ \frac{X - 3.35}{0.15} > z_{0.025} \right\} \\
&= \left\{ \, X < 3.35 - 0.15 \times 1.96 \right\} \cup \left\{ \, X > 3.35 + 0.15 \times 1.96 \right\} \\
&= \left\{ \, X < 3.056 \right\} \cup \left\{ \, X > 3.644 \right\}
\end{aligned}
$$

(b) We want the probability of rejecting the null when H_a is true (specifically, when $\mu = 3.40$). With 100 loads of milk sampled, $\bar{X} \sim N(\mu, \sigma/\sqrt{100})$.

First, the rejection region for the mean of the 100 samples under the null is:

$$
\begin{aligned}
R &= \left\{ \frac{\bar{X} - 3.35}{0.015} < -z_{0.025} \right\} \cup \left\{ \frac{\bar{X} - 3.35}{0.015} > z_{0.025} \right\} \\
&= \left\{ \, \bar{X} < 3.35 - 0.015 \times 1.96 \right\} \cup \left\{ \, \bar{X} > 3.35 + 0.015 \times 1.96 \right\} \\
&= \left\{ \, \bar{X} < 3.3206 \right\} \cup \left\{ \, \bar{X} > 3.3794 \right\}
\end{aligned}
$$

We can now compute the probability of rejecting the null when in fact $\mu = 3.40$:

$$
\begin{aligned}
1 - \beta &= P(\, \bar{X} < 3.3206 \mid \mu = 3.40) + P(\, \bar{X} > 3.3794 \mid \mu = 3.40) \\
&= P\left(\frac{\bar{X} - 3.40}{0.015} < \frac{3.3206 - 3.40}{0.015} \right) + P\left(\frac{\bar{X} - 3.40}{0.015} > \frac{3.3794 - 3.40}{0.015} \right) \\
&= P(\, Z < -5.293333) + P(\, Z > -1.373333) \\
&= \texttt{(negligible quantity)} + 1 - \texttt{pnorm(-1.373333)} \\
&= 0.9151756
\end{aligned}
$$

$$
\boxed{\text{power} = 1 - \beta = 0.915}
$$

(c) We now we express the power in terms of the change of the mean over the standard deviation ($|\mu_1 - \mu_0|/\sigma$) and plot the result.

$$
\begin{aligned}
1 - \beta &= P\left(\left| \frac{\bar{X} - \mu_1 + (\mu_1 - \mu_0)}{\sigma} \right| < z_{0.025} \right) \\
&= P\left(\left| Z + \frac{\mu_1 - \mu_0}{\sigma} \right| < 1.96 \right) \\
&= P\left(Z > -1.96 - \frac{|\mu_1 - \mu_0|}{\sigma} \right) + P\left(Z < 1.96 - \frac{|\mu_1 - \mu_0|}{\sigma} \right)
\end{aligned}
$$

2

(d)

Question 7.2. The relative rotation angle between the **L2** and **L3** lumbar vertebrate is defined as the acute angle between posterior tangents drawn to each vertebra on a spinal **X**-ray. When this angle is too large the patient experiences discomfort or pain. Chiropractic treatment of this condition involves decreasing this angle by applying (nonsurgical) manipulation or pressure. Harrison et al. (2002) propose one such particular treatment. They measured the angle on both pre- and post-treatment **X**-rays from a random sample of 48 patients. At $\alpha = 0.05$, test whether the mean post-treatment angle is less than the mean angle prior to treatment.

You can load the data using the following command:

```
har <- read.table("http://astro.temple.edu/ zhaozhg/Stat8003/data/har1.cs
sep=",", header=TRUE)
```

Answer:

(a)

Question 7.3. This problem will guide you to demonstrate the multiplicity issues in multiple hypothesis testing using simulation. Let X_{ij} be the data modeled as

$$X_{ij} \overset{iid}{\sim} N(\theta_j, 1), \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, p. \tag{1}$$

Consider testing p hypotheses $H_{0j} : \theta_j = 0 \ vs H_{1j} : \theta_j \neq 0$ where $j = 1, 2, \ldots, p$. Define the family-wise error rate (FWER) as

$$FWER = P \text{ (at least one (including one) false rejection)}.$$

Set the sample size $n = 50$ and $\alpha = 0.05$.

(a) Set $p = 1$, $\theta_j = 0$, $\forall j$. Generate X_{ij} according to (1) and use p-value to test the hypothesis at α-level. Replicate this step 1000 times to get the simulated FWER.

(b) Set $p = 10$, $\theta_j = 0$, $\forall j$. Generate X_{ij} according to (1) and use p-value to test 10 hypothesis at α-level. Replicate this step 1000 times to get the simulated FWER.

(c) Set $p = 100$, $\theta_j = 0$, $\forall j$. Generate X_{ij} according to (1) and use p-value to test 100 hypothesis at α-level. Replicate this step 1000 times to get the simulated FWER.

(d) Set $p = 100$, $\theta_j = 0$, $\forall j$. Generate X_{ij} according to (1) and use p-value to test 100 hypothesis using Bonferroni's correction by setting the significance level at α/p for each hypothesis. Replicate this step 1000 times to get the simulated FWER.

Answer:

(a)