Statistics for Data Science - 2

Graded assignment week 10

Use the following values of standard normal distribution if needed.

$F_Z(0.15) = 0.55962, F_Z(-0.04) = 0.48405, F_Z(-1.28) = 0.10027, F_Z(1.96) = 0.975, F_Z(-1.64) = 0.05, F_Z(1.64) = 0.95, F_Z(-1.28) = 0.01, F_Z(2.74) = .99693, F_Z(-1.36) = .08691, F_Z(1.28) = .89973, F_Z(2.54) = .99446, F_Z(-2.8) = .00256$

1. The average marks scored by students of a school in their board exams is reported to be 400 with a standard deviation of 5. You suspect that the average may be lower, possibly 390, and decide to sample students to find their marks.

   (a) What sample size do you need for a test at the significance level 0.05 and power 0.95?
   Answer: 3

   Solution:
   Let the random variable X represent the marks obtained by students in their board exams with expected value μ and standard deviation σ.
   Given $\mu = 400$ and $\sigma = 5$.
   Consider the Null and alternative hypothesis:

$$H_0 : \mu = 400$$

$$H_A : \mu < 400$$

   Test Statistics: $\overline{X}$
   Test: Reject H_0, if $\overline{X} < c$ at $\alpha = 0.05$

$$\begin{aligned}
\alpha &= P(\text{reject } H_0 | H_0 \text{ is true}) \\
&= P\left(\frac{\overline{X} - 400}{5/\sqrt{n}} < \frac{c - 400}{5/\sqrt{n}}\right) \\
&\implies 0.05 = F_Z\left(\frac{c - 400}{5/\sqrt{n}}\right) \\
&\implies F_Z^{-1}(0.05) = \frac{c - 400}{5/\sqrt{n}} \\
&\implies -1.64 = \frac{c - 400}{5/\sqrt{n}} \\
&\implies c = -1.64 \times \frac{5}{\sqrt{n}} + 400 \qquad \cdots (1)
\end{aligned}$$

Again, when alternative hypothesis is true, we have

$$\frac{\overline{X}-390}{5/\sqrt{n}} \sim \text{Normal}(0,1)$$

$$\begin{aligned}
&\beta = P(\text{Accept } H_0 | H_A \text{ is true})\\
&P(\overline{X} \geq c \mid \mu = 390)\\
&P\left(\frac{\overline{X}-390}{5/\sqrt{n}} \geq \frac{c-390}{5/\sqrt{n}}\right)\\
&\implies 0.05 = 1 - F_Z\left(\frac{c-390}{5/\sqrt{n}}\right)\\
&\implies F_Z^{-1}(0.95) = \frac{c-390}{5/\sqrt{n}}\\
&\implies 1.64 = \frac{c-390}{5/\sqrt{n}}\\
&\implies c = 1.64 \times \frac{5}{\sqrt{n}} + 390 \qquad \cdots (2)
\end{aligned}$$

From equation (1) and (2), we have

$$\begin{aligned}
&-1.64 \times \frac{5}{\sqrt{n}} + 400 = 1.64 \times \frac{5}{\sqrt{n}} + 390\\
&\implies 2 \times 1.64 \times \frac{5}{\sqrt{n}} = 10\\
&\implies n = 1.64^2 = 2.6896 \approx 3
\end{aligned}$$

(b) Find the critical value c. Enter the answer correct to two decimal places.
394.73, [394.70, 395.30]
Solution: Substituting the value of c in (2), we get $c = 394.73$.

2. Suppose $X \sim \text{Normal}(\mu, 9)$. For $n = 100$ iid samples of X, the observed sample mean is 11.8. What conclusion would a z-test reach if the null hypothesis assumes $\mu = 10.5$ (against an alternative hypothesis $\mu \neq 10.5$)?

(a) Accept H_0 at a significance level of 0.10.

(b) Reject H_0 at a significance level of 0.10.

(c) Accept H_0 at a significance level of 0.05.

(d) Reject H_0 at a significance level of 0.05.

Solution:
Given, $X \sim \text{Normal}(\mu, 9)$.

$X_1, \ldots, X_{100} \sim iid\ X$. For 100 iid samples of X, $\bar{X} \sim \text{Normal}(\mu, 9/100)$
Sample mean, $\bar{X} = 11.8$.
Consider the null and alternative hypotheses: Null hypothesis,

$$H_0 : \mu = 10.5$$

$$H_A : \mu \neq 10.5$$

Test for $\alpha = 0.05$
Test: Reject H_0 if $|\bar{X} - \mu| > c$

$$\alpha = P(\mid \bar{X} - 10.5 \mid > c | \mu = 10.5)$$

$$\Rightarrow \alpha = P\left(\mid \frac{\bar{X} - 10.5}{\sqrt{9/100}} \mid > \frac{c}{\sqrt{9/100}}\right)$$

$$\Rightarrow \alpha = P\left(\mid z \mid > \frac{c}{\sqrt{9/100}}\right)$$

$$\Rightarrow 0.05 = 2F_z\left(\frac{-c}{\sqrt{9/100}}\right)$$

$$\Longrightarrow \ c = -\frac{3}{10} F_z^{-1}(0.025) = 0.5879 \tag{1}$$

Since $\mid 11.8 - 10.5 \mid > c$, z-test at significance level $(\alpha) = 0.05$, will reject H_0.

Test for $\alpha = 0.1$
Test: Reject H_0 if $|\bar{X} - \mu| > c$

$$\alpha = P(\mid \bar{X} - 10.5 \mid > c | \mu = 10.5)$$

$$\alpha = P\left(\mid \frac{\bar{X} - 10.5}{\sqrt{25/100}} \mid > \frac{c}{\sqrt{9/100}}\right)$$

$$\alpha = P\left(\mid z \mid > \frac{c}{\sqrt{9/100}}\right)$$

$$0.1 = 2F_z\left(\frac{-c}{\sqrt{9/100}}\right)$$

$$c = -\frac{3}{10} F_z^{-1}(0.05) = 0.4934 \tag{2}$$

Since $\mid 11.8 - 10.5 \mid > c$, z-test at significance level $(\alpha) = 0.10$, will reject H_0.
Hence, options (b) and (d) are correct.

3. Let $X_1, \ldots, X_{100}$ be a sample from a normal distribution having a variance of 25. We wish to test the hypothesis $H_0 : \mu = 0$ versus $H_A : \mu = 1.5$. Consider a test that rejects H_0 for $\bar{X} > c$.

   (a) Find the value of c at a significance level $\alpha = 0.05$. Enter the answer correct to two decimal places.
   0.82
   Solution:
   Given, $X \sim \text{Normal}(\mu, 25)$.

   $X_1, \ldots, X_{100} \sim iid\ X$. For 100 iid samples of X, $\bar{X} \sim \text{Normal}(\mu, 25/100)$
   The null and alternative hypothesis are

$$H_0 : \mu = 0$$

$$H_A : \mu > 0$$

   Test for $\alpha = 0.05$
   Test: Reject H_0 if $\bar{X} > c$

$$\alpha = P(\bar{X} > c | \mu = 0)$$
$$\Rightarrow \alpha = P\left(\frac{\bar{X}}{\sqrt{25/100}} > \frac{c}{\sqrt{25/100}}\right)$$
$$\Rightarrow \alpha = P\left(z > 2c\right)$$
$$\Rightarrow 0.05 = 1 - F_z\left(2c\right)$$
$$\Rightarrow c = \frac{1}{2} F_z^{-1}(0.95) = 0.8224$$

   (b) Find the power of the test. Enter the answer correct to two decimal places.
   0.91309, [0.90, 0.93]

Solution:

$$\text{Power} = 1 - \beta = P(\bar{X} > c | \mu = 1.5)$$
$$\Rightarrow 1 - \beta = P\left(\frac{\overline{X} - 1.5}{\sqrt{25/100}} > \frac{c - 1.5}{\sqrt{25/100}}\right)$$
$$\Rightarrow 1 - \beta = P\left(z > \frac{c - 1.5}{\sqrt{25/100}}\right)$$
$$\Rightarrow 1 - \beta = 1 - F_z\left(2(c - 1.5)\right)$$

Substituting the value of c from above problem,

$$1 - \beta = 1 - F_z\left(2(0.8224 - 1.5)\right)$$

Therefore,

$$\text{Power} = 1 - \beta = 0.9123$$

4. A manufacturer supplies fuses, approximately 90% of which function properly. A new process is initiated whose purpose is to increase the proportion of properly functioning fuses. We obtain a random sample of 100 such fuses manufactured by the new process and found out that 8 of them are not functioning properly. Let p denotes the proportion of properly functioning fuses. (Use normal approximation to binomial)

   (a) Define null hypothesis and alternative hypothesis.

      i. $H_0 : p = 0.90,\ H_A : p \neq 0.90$
      ii. $H_0 : p = 0.90,\ H_A : p < 0.90$
      iii. $H_0 : p = 0.90,\ H_A : p > 0.90$
      iv. $H_0 : \overline{X} = 0.90,\ H_A : \overline{X} > 0.90$

   (b) Choose the correct options from the following:

      i. Accept H_0 at a significance level of 0.05.
      ii. Reject H_0 at a significance level of 0.10.
      iii. Accept H_0 at a significance level of 0.10.
      iv. Reject H_0 at a significance level of 0.05.

Solution:
$X_1, \ldots, X_{100} \sim$ iid Bernoulli(p).

The null and alternative hypothesis are:

$$H_0 : p = 0.9$$

$$H_A : p > 0.9$$

Sample mean $(\bar{X}) = \dfrac{92}{100} = 0.92$
Test for $\alpha = 0.05$
Test statistic, $T = X_1 + \ldots + X_{100} \sim$ Binomial(n, p) which can be normally approximated as
$T \approx \text{Normal}(100p, 100p(1-p))$
$\overline{X} \approx \text{Normal}\left(p, \dfrac{p(1-p)}{100}\right)$
Test: Reject H_0 if $\bar{X} > c$

$$\alpha = P(\bar{X} > c \mid p = 0.9)$$

$$\alpha = P\left(\frac{\bar{X} - p}{\sqrt{\dfrac{p(1-p)}{n}}} > \frac{c - p}{\sqrt{\dfrac{p(1-p)}{n}}}\right)$$

$$\alpha = P\left(z > \frac{c - 0.9}{\sqrt{\dfrac{0.9 \times 0.1}{100}}}\right)$$

$$\Rightarrow 0.05 = 1 - F_z\left(\frac{c - 0.9}{0.03}\right)$$

$$\Rightarrow c = 0.9 + 0.03 \times F_z^{-1}(0.95)$$

$$c = 0.9493$$

Since $\bar{X} < c$, z-test at significance level $(\alpha) = 0.05$, will accept H_0.
Test for $\alpha = 0.10$
Test statistic, $T = X_1 + \ldots + X_{100} \sim \text{Binomial}(n, p)$ which can be normally approximated as
$T \approx \text{Normal}(100p, 100p(1-p))$

$\bar{X} \approx \text{Normal}\left(p, \dfrac{p(1-p)}{100}\right)$

Test: Reject H_0 if $\bar{X} > c$

$$\alpha = P(\bar{X} > c \mid p = 0.9)$$

$$\alpha = P\left(\frac{\bar{X} - p}{\sqrt{\dfrac{p(1-p)}{n}}} > \frac{c - p}{\sqrt{\dfrac{p(1-p)}{n}}}\right)$$

$$\alpha = P\left(z > \frac{c - 0.9}{\sqrt{\dfrac{0.9 \times 0.1}{100}}}\right)$$

$$\Rightarrow 0.10 = 1 - F_z\left(\frac{c - 0.9}{0.03}\right)$$

$$\Rightarrow c = 0.9 + 0.03 \times F_z^{-1}(0.90)$$

$$\Rightarrow c = 0.9384$$

Since $\bar{X} < c$, z-test at significance level $(\alpha) = 0.10$, we will accept H_0.
Hence, options (i) and (iii) are correct.

5. A commonly prescribed drug for relieving nervous tension is believed to be only 25% effective. To determine if a new drug is superior in providing relief, suppose that 100 people who were suffering with nervous tension are chosen at random and inoculated. If more than 36 of them are found to be relieved, we reject the null hypothesis that $p = 1/4$ and the new drug will be considered superior to the one presently in use. (Use

normal approximation to binomial)

(a) Find the critical value c.
Answer: 36

Solution: Since, we will reject the null hypothesis if more than 36 out of 100 patients is found to be relieved, 36 is the critical value.

(b) Find $P(\text{Type I error})$. Enter the answer correct to four decimal places.
[0.0038, 0.0060]
Solution:

The null and alternative hypothesis are:

$$H_0 : p = 0.25$$

$$H_A : p > 0.25$$

Given, critical value (c) $= 36$
Test statistic, $T = X_1 + \ldots + X_{100} \sim \text{Binomial}(100, p)$ which can be normally approximated as
$T \approx \text{Normal}(100p, 100p(1-p))$

Test: Reject H_0 if $T > c$ that is $T > 36$

$$\alpha = P(T > c \mid p = 0.25)$$

$$\alpha = P(T > 36 \mid p = 0.25)$$

$$\alpha = P\left(z > \frac{36 - 100p}{\sqrt{100p(1-p)}}\right)$$

$$\alpha = P\left(z > \frac{36 - 100(0.25)}{\sqrt{100 \times 0.25(0.75)}}\right)$$

$$\alpha = P\left(z > \frac{11}{\sqrt{18.75}}\right)$$

$$\alpha = 1 - F_z\left(\frac{11}{\sqrt{18.75}}\right)$$

$$\alpha = 1 - F_z(2.54)$$

$$P(\text{Type I error}) = \alpha = 0.0055$$

(c) Find $P(\text{Type II error})$ for $p = 1/2$. Enter the answer correct to four decimal places.
[0.0024, 0.0035]

Solution:

$$P(\text{Type II error}) = \beta = P(T \leq c \sim p = 0.5)$$

$$\beta = P(T \leq 36 \sim p = 0.5)$$

$$\beta = P\left(z \leq \frac{36 - 100p}{\sqrt{100p(1-p)}}\right)$$

$$\beta = P\left(z \leq \frac{36 - 100(0.5)}{\sqrt{100 \times 0.5(0.5)}}\right)$$

$$\beta = P\left(z \leq \frac{-14}{\sqrt{25}}\right)$$

$$\beta = F_z\left(\frac{-14}{5}\right)$$

$$\beta = F_z(-2.8)$$

$$P(\text{Type II error}) = \beta = 0.0025$$

6. The proportion of adults living in a small town who are college graduates is estimated to be $p = 0.6$. To test this hypothesis against the alternative $p < 0.6$, you decide to take a sample of adults from the town.

(a) What sample size do you need for a test (against the alternative hypothesis that $p = 0.4$) at a significance level of 0.10 and power of 0.90?
40

Solution:

Null hypothesis, $H_0 : p = 0.6$

Alternate hypothesis, $H_A : p < 0.6$

Given, $\alpha = 0.10$ and power $1 - \beta = 0.90$

Test statistic, $T = \text{Binomial}(n, p)$ which can be normally approximated as

$T \approx \text{Normal}(np, np(1-p))$

$\overline{X} \approx \text{Normal}\left(p, \dfrac{p(1-p)}{n}\right)$

Test: Reject H_0 if $T < c$

$$\alpha = P(\bar{X} < c | p = 0.4)$$

$$\alpha = P\left(\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{c - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\alpha = P\left(z < \frac{c - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$\alpha = P\left(z < \frac{c - 0.6}{\sqrt{\frac{0.6 \times 0.4}{n}}}\right)$$

$$\Rightarrow 0.10 = F_z\left(\frac{c - 0.6}{\sqrt{\frac{0.24}{n}}}\right)$$

$$\Rightarrow c = 0.6 + \sqrt{\frac{0.24}{n}} F_z^{-1}(0.10)$$

$$c = 0.6 - \frac{0.6278}{\sqrt{n}} \tag{3}$$

Now, power

$$1 - \beta = P(\bar{X} < c \mid p = 0.6)$$

$$1 - \beta = P\left(\frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} < \frac{c - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$1 - \beta = P\left(z < \frac{c - p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

$$1 - \beta = P\left(z < \frac{c - 0.4}{\sqrt{\frac{0.4 \times 0.6}{n}}}\right)$$

$$\Rightarrow 0.90 = F_z\left(\frac{c - 0.4}{\sqrt{\frac{0.24}{n}}}\right)$$

$$\Rightarrow c = 0.4 + \sqrt{\frac{0.24}{n}} F_z^{-1}(0.90)$$
$$c = 0.4 + \frac{0.6278}{\sqrt{n}} \tag{4}$$

Solving equations (3) and (4),

$$n \approx 39.41$$
$$n = 40$$

(b) Find the critical value at a significance level of 0.10. Enter the answer correct to two decimal places.
[0.48, 0.52]
Solution:
Substitute $n = 40$ in equation (3),

$$c = 0.5$$

7. A random sample of 36 packets of marshmallow weighs, on average, 145 grams with a standard deviation of 5 grams. Test the hypothesis that $\mu = 150$ grams against the alternative hypothesis, $\mu < 150$ grams, at the 0.05 level of significance.

   (a) On average, it weighs less than 150 grams.
   (b) On average, it weighs 150 grams.

   Solution:

   The null and the alternative hypothesis are:

$$H_0 : \mu = 150$$
$$H_A : \mu < 150$$

   Test: Reject H_0, if $\overline{X} < c$.
   Given $\alpha = 0.05$, we have

$$\alpha = P(\overline{X} < c \mid \mu = 150)$$
$$\implies 0.05 = P\left(\frac{\overline{X} - 150}{5/\sqrt{36}} < \frac{c - 150}{5/\sqrt{36}}\right)$$
$$\implies 0.05 = F_Z\left(\frac{c - 150}{5/\sqrt{36}}\right)$$
$$\implies -1.64 = \frac{c - 150}{5/\sqrt{36}} \implies c = 148.63$$

   Since $\overline{X} = 145 < c$, reject H_0.

8. A survey of 225 randomly selected students from a city revealed that 89.4% of them have participated in extra curricular activities in their schools. Can we conclude at 1% level of significance that 90% of the students have participated in extra curricular activities?

   (a) Yes
   (b) No

   Solution:
   Null hypothesis, $H_0 : p = 0.9$

   Alternate hypothesis, $H_A : p \neq 0.9$

   Given, $\alpha = 0.1$ and $n = 225$

$$\alpha = P(\mid \bar{X} - p \mid > c \mid \mu = 0.9)$$

$$\alpha = P\left(\mid \frac{\bar{X} - 0.9}{\sqrt{\dfrac{0.9 \times 0.1}{225}}} \mid > \frac{c}{\sqrt{\dfrac{0.9 \times 0.1}{225}}} \right)$$

$$\alpha = P\left(\mid z \mid > \frac{c}{\sqrt{\dfrac{0.9 \times 0.1}{225}}} \right)$$

$$0.1 = 2F_z\left(\frac{-15c}{\sqrt{0.9 \times 0.1}} \right)$$

$$c = -\frac{\sqrt{0.9 \times 0.1}}{15} \times F_z^{-1}(0.05) = 0.03289$$

   Since $\mid \bar{X} - p \mid = \mid 0.894 - 0.9 \mid < c$, z-test at significance level $(\alpha) = 0.05$, will accept H_0.

9. A box of a certain brand of washing powder advertises that it weighs 2.5 kg, but the actual weight is 2.4 kg with a standard deviation of 0.1 kg. The company wants to test if the mean has changed. They take a random sample of 100 boxes and finds that the average weight is 2.35 kg.

   (a) Define null hypothesis and alternative hypothesis.

      i. $H_0 : \mu = 2.4, \;\; H_A : \mu \neq 2.4$
      ii. $H_0 : \mu = 2.4, \;\; H_A : \mu > 2.4$
      iii. $H_0 : \mu = 2.4, \;\; H_A : \mu < 2.4$

iv. $H_0 : \mu = 2.5,\ H_A : \mu \neq 2.5$

(b) What conclusion should be made using a significance level of $\alpha = 0.05$?

i. Accept H_0.

ii. Reject H_0 and accept H_A.

Solution:

The company wants to check if the mean has changed. So, null and alternative hypothesis are given by

$$H_0 : \mu = 2.4, \quad \mu \neq 2.4$$

Define a test statistic T as $T = \overline{X}$.

Test: reject the null hypothesis if $|\overline{X} - 2.4| > c$.

By CLT, we can say that $\dfrac{\overline{X} - 2.4}{0.1/\sqrt{100}} = \dfrac{\overline{X} - 2.4}{1/100} \sim \text{Normal}(0, 1)$.

Now,

$$\begin{aligned}
\alpha &= P(|\overline{X} - 2.4| > c) \\
\implies 0.05 &= P\left(\left|\frac{\overline{X} - 2.4}{1/100}\right| > \frac{c}{1/100}\right) \\
\implies 0.05 &= P(|Z| > 100c) \\
\implies 0.05/2 &= P(Z < -100c) \\
\implies -1.96 &= -100c \implies c = 0.0196
\end{aligned}$$

Since $|\overline{X} - 2.4| = |2.35 - 2.4| = 0.05 > c$, reject H_0.

10. It is claimed that the lifetimes of light bulbs are normally distributed with a mean of 800 hours and a standard deviation of 40 hours. We wish to test the hypothesis that $\mu = 800$ hours against the alternative that $\mu \neq 800$ hours with a sample size of 30.

(a) If the acceptance region is defined as $780 \leq \bar{X} \leq 820$, find the significance level. Enter the answer correct to three decimal places.

0.006, [0.005, 0.008]

Solution:

Let the random variable X denote the lifetime of electric bulbs.

Given, $X \sim \text{Normal}(\mu, 40^2)$.

$X_1, \ldots, X_{30} \sim iid\ X$.

For 30 iid samples of X, $\bar{X} \sim \text{Normal}(\mu, 40^2/30)$

Null hypothesis, $H_0 : \mu = 800$

Alternate hypothesis, $H_A : \mu \neq 800$

$$\alpha = P(\text{Reject } H_0 \mid H_0 \text{ is true})$$
$$\alpha = P(\bar{X} > 820 \text{ or } \bar{X} < 780 \mid \mu = 800)$$
$$\alpha = P(\mid \bar{X} - 800 \mid > 20)$$
$$\alpha = P\left(\mid \frac{\bar{X} - 800}{\sqrt{1600/30}} \mid > \frac{20}{\sqrt{1600/30}}\right)$$
$$\alpha = P\left(\mid z \mid > \frac{20}{\sqrt{1600/30}}\right)$$
$$\alpha = 2F_z\left(\frac{-20}{\sqrt{1600/30}}\right)$$
$$\alpha = 0.006$$

(b) Find the power of the test against the alternative that if the true mean life is 788 hours. Enter the answer correct to two decimal places.
[0.91, 0.94]
Solution:

$$\text{Power} = 1 - \beta = P(\text{Reject } H_0 \mid H_A \text{ is true})$$
$$1 - \beta = P(\bar{X} > 820 \text{ or } \bar{X} < 780 \mid \mu = 788)$$
$$1 - \beta = P(\bar{X} > 820 \mid \mu = 788) + P(\bar{X} < 780 \mid \mu = 788)$$
$$1 - \beta = P\left(\frac{\bar{X} - 788}{\sqrt{1600/30}} > \frac{820 - 788}{\sqrt{1600/30}}\right) + P\left(\frac{\bar{X} - 788}{\sqrt{1600/30}} < \frac{780 - 788}{\sqrt{1600/30}}\right)$$
$$1 - \beta = P\left(z > \frac{32}{\sqrt{1600/30}}\right) + P\left(z < \frac{-8}{\sqrt{1600/30}}\right)$$
$$1 - \beta = 1 - F_z\left(\frac{32}{\sqrt{1600/30}}\right) + F_z\left(\frac{-8}{\sqrt{1600/30}}\right)$$
$$1 - \beta = 0.1366$$