BSCMA1004

STATISTICS II NOTES

WEEK 11 NOTES

IITM B.S Degree

PREPARED BY

Vehaan Handa, IIT Madras

Stats - II - Week XI Notes

Hypothesis Testing

An authentic coin or a counterfeit coin?
It is known that authentic $\Rightarrow P(H) = 0.5$, counterfeit $\Rightarrow P(H) = 0.6$
How will you test for the authenticity? You may toss the coin multiple times, observe results.

- Decide between a null hypothesis H_0 and alternative hypothesis H_A, using samples.

Here, $H_0 : P(H) = 0.5$ and ~~X~~ $H_A : P(H) = 0.6$.
So you either accept or reject H_0.

Suppose coin tossed 3 times :→
- Possible outcomes : HHH, HHT, ..., TTT.
- For some outcomes, accept H_0. For others, reject.
- ~~X~~ If A is the subset of outcomes for which we accept H_0, every acceptance subset A corresponds to a test.

$X_1, \dots, X_n \sim \text{iid } X$, H_0 : null hypothesis H_A : alternative
- Suppose $X \in \mathcal{X}$. Samples $X_1, \dots, X_n \in \mathcal{X}^n$
- Subset $A \subseteq \mathcal{X}^n \leftrightarrow$ hypothesis test

If $X_1, \dots, X_n \in A$, we accept H_0, otherwise reject H_0.

3 coin tosses, 8 outcomes
$2^8 = 256$ subsets $\leftrightarrow$ 256 tests
How to define a good acceptance set/ test?

Metric 1: Significance Level (size) of test, denoted α.

- Type I Error: Rejected H_0 when H_0 is true
- $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true})$

Metric 2: Power of a Test, $1-\beta$

- Type II Error: Accept H_0 when H_A is true
- $\beta = P(\text{Type II error}) = P(\text{Accept } H_0 \mid H_A \text{ is true})$
- Power $= 1-\beta = P(\text{Reject } H_0 \mid H_A \text{ is true})$

Computing α, β

$H_0 : P(H) = 0.5$, $H_A : P(H) = 0.6$
Toss 3 times. $\mathcal{X}^3 = \{HHH, \ldots, TTT\}$

Case
- $A = \emptyset$ (don't accept H_0 at all)
- $\alpha = 1$, $\beta = 0$

Case
- $A = \mathcal{X}^3$ (always accept H_0)
- $\alpha = 0$, $\beta = 1$

Case
- $A = \{HHT, HTH, HTT, THH, THT, TTH\}$
- $\alpha = P(A^c \mid P(H) = 0.5) = 2/8 = 0.25$
- $\beta = P(A \mid P(H) = 0.6) = \underbrace{3(0.4)^2(0.6)}_{2T, 1H} + \underbrace{3(0.4)(0.6)^2}_{1T, 2H}$

$= 0.72$

Case
- $A = \{TTT, TTH, THT, HTT\}$
- $\alpha = 4/8 = 0.5$
- $\beta = \underbrace{0.4^3}_{3T} + \underbrace{3(0.4)^2(0.6)}_{2T, 1H} = 0.352$

DATE

Interesting thing: If you plot β vs α for all 256 tests, and if you fix a particular α on x axis, you get a least β value.
But not always feasible. What if we toss 100 times? And what about other distributions?
But theoretically, this is a nice thing to think about.
In hypothesis testing, you pick a significance level first, and then hope that you have as high of a power as possible.

☆ Neyman-Pearson Paradigm of Hypothesis Testing

$X_1, \ldots, X_n \sim$ iid X

- H_0 : null hypothesis. on dist. of X H_A: alternative hypothesis.

Test is defined by acceptance set A. If samples fall in A, accept H_0, otherwise reject H_0.

- Two errors and two metrics defined. earlier.

Q. An establishment claims a spot remover will remove >80% of spots. Spot remover will be used on 10 spots at random. If fewer than 8 are removed, then we shall not reject $H_0: p = 0.8$, else we conclude $H_A: p > 0.8$. Find α. Find β for the alternative that $p = 0.7$.

$\alpha = P(\text{reject } H_0 / H_0) \rightarrow$ we reject null when ≥ 8 spots are removed

$$\alpha = {}^{10}C_8 (0.8)^8 (0.2)^2 + {}^{10}C_9 (0.8)^9 (0.2) + {}^{10}C_{10} (0.8)^{10} = \underline{\underline{0.67}}$$

$$1-\beta = {}^{10}C_8 (0.7)^8 (0.3)^2 + {}^{10}C_9 (0.7)^9 (0.3) + {}^{10}C_{10} (0.7)^{10} = 0.37$$

$$\boxed{\beta = 0.63}$$

Q. Consider 100 coin tosses, either authentic / counterfeit $(P(H)=0.5)$ $(P(H)=0.6)$
Suppose T = no. of heads seen. Consider a test that rejects H_0 if $T > c$ for some c. What's the significance level? What's the power?

T = no. of heads . $A = \{\text{outcomes} : T \le c\}$

$T \sim Bin(100, P(H))$

$$\alpha = P(\underbrace{\text{Reject } H_0}_{\text{outcome} \in A^c} \mid \underbrace{H_0}_{P(H)=0.5})$$

$$= P(A^c \mid P(H) = 1/2) = \sum_{k=c+1}^{100} \binom{100}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{100-k}$$

Fix α and find c.

$$1-\beta = P(\underbrace{\text{Reject } H_0}_{\text{outcome} \in A^c} \mid \underbrace{H_A}_{P(H)=0.6})$$

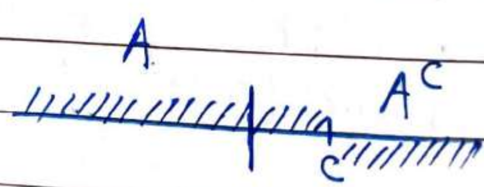
$$= P(A^c \mid P(H) = 0.6) = \sum_{k=c+1}^{100} \binom{100}{k} 0.6^k \, 0.4^{100-k}$$

We want to maximise this power, so we have to keep c as low as possible. Lower $c \Rightarrow$ Higher Power

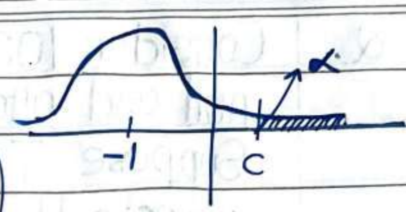
→ Fix α. Find lowest possible c.

Q. Consider sample $X \sim Normal(\mu, 1)$. Let null and alternative hypothesis be $H_0 : \mu = -1$ and $H_A : \mu = 1$. Consider a test that rejects H_0 if $X > c$ for some c. Significance level? Power?

$A = \{X \le c\}$



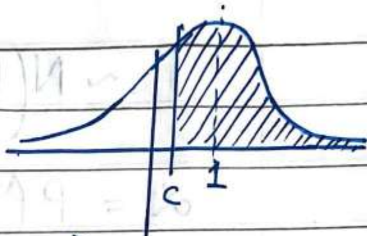
$$\alpha = P(A^c \mid \mu = -1) = P(N(-1,1) > c)$$



$$= P\left(\frac{N(-1,1)+1}{1} > \frac{c+1}{1}\right)$$

$$= P(Z > c+1) = 1 - F_Z(c+1)$$

$$1-\beta = P(A^c \mid \mu = 1) = P(N(1,1) > c)$$

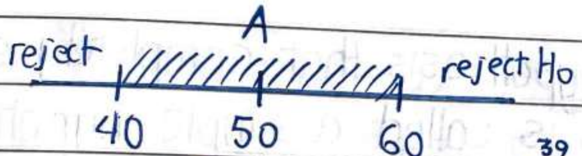


$$= P\left(\frac{N(1,1)-1}{1} > c-1\right)$$

$$= P(Z > c-1) = 1 - F_Z(c-1)$$

Pick a low value of c.

Q. Consider $X \sim \text{Binomial}(100, p)$. Let null and alternative hypothesis be: $H_0: p = 0.5$, $H_A: p \neq 0.5$. Consider a test that rejects H_0 if $|X-50| > 10$. Significance level? Power as fn. of p? Use normal approximation.



$A = \{40, 41, \ldots, 60\}$

$$\alpha = P(A^c \mid p = 0.5) = \sum_{k=0}^{39} \binom{100}{k}\left(\frac{1}{2}\right)^{100} + \sum_{k=61}^{100} \binom{100}{k}\left(\frac{1}{2}\right)^{100}$$

$$1-\beta = P(A^c \mid p) = P(|X-50| > 10 \mid p) = P(X > 60 \mid p) + P(X < 40 \mid p)$$

$X \sim \text{Normal}(\text{mean} = 100p,\ \text{Variance} = 100p(1-p))$

$$\approx P\left(\frac{X-100p}{\sqrt{100p(1-p)}} > \frac{60-100p}{\sqrt{100p(1-p)}}\right) + P\left(\frac{X-100p}{\sqrt{100p(1-p)}} < \frac{40-100p}{\sqrt{100p(1-p)}}\right)$$

$$\approx P\left(Z > \frac{6-10p}{\sqrt{p(1-p)}}\right) + P\left(Z < \frac{4-10p}{\sqrt{p(1-p)}}\right)$$

$$= 1 - F_Z\left(\frac{6-10p}{\sqrt{p(1-p)}}\right) + F_Z\left(\frac{4-10p}{\sqrt{p(1-p)}}\right)$$

DATE

Q. Consider 100 samples $X_1, \ldots, X_{100} \sim N(\mu, 1)$. Let the null and alternative hypothesis be $H_0 : \mu = -1$, $H_A : \mu = 1$. Suppose $T = (X_1 + \ldots + X_{100})/100$. Consider a test that rejects H_0 if $T > c$ for some c. Significance level? Power?

$T \sim N\left(\mu, \frac{1}{100}\right) \qquad A = \{T \leq c\}$

$\alpha = P(T > c \mid \mu = -1) = P\left(Z > \frac{c+1}{1/10}\right) = P(Z > 10(c+1))$

$= 1 - F_Z(10(c+1))$

$1-\beta = P(T > c \mid \mu = 1) = P(Z > 10(c-1)) = 1 - F_Z(10(c-1))$

★ Types of Hypothesis Testing

① Simple Hypothesis : A hypothesis that completely specifies distribution of samples is called a simple hypothesis.

Eg. P(heads) = 0.5 , P(Heads) = 0.9
Eg. Normal $(\mu, 3) \rightarrow \mu = 1, \mu = -1$ etc.

Simple Null vs. Simple Alternative :→
Very well understood, best approach known. Rarely occurs.

② Composite Hypothesis : Hypothesis that does not completely specify distribution of samples.

Eg. Coin Toss: Null : P(H) = 0.5 , fair coin, simple.
Alt : P(H) ≠ 0.5, unfair coin, Composite.

Normal $(\mu, 3)$: Null : $\mu = 0$ (some effect not present), simple
$\mu > 1$ (some effect present) composite.

Simple/Composite Null vs alternative ....

DATE

Standard Tests for One Sample

$X_1 \ldots, X_n \sim iid\ X$, $E[X] = \mu$, $Var(X) = \sigma^2$

Testing for mean, null $H_0 : \mu = c$

- Alternative :→ Right tail test, $H_A : \mu > c$
  Left tail test, $H_A : \mu < c$
  Two tail test, $H_A : \mu \neq c$

Two cases : known or unknown variance

Testing for ~~the~~ variance :→ Null $H_0 : \sigma = c$
Alt. $H_A : \sigma > c$

Two Samples

$X_1 \ldots, X_{n_1} \sim iid\ X$, $E[X] = \mu_1$, $Var(X) = \sigma_1^2$
$Y_1, \ldots, Y_{n_2} \sim iid\ Y$, $E[Y] = \mu_2$, $Var(X) = \sigma_2^2$

$$\left.\begin{array}{l} \text{Null } H_0 : \mu_1 = \mu_2 \\ H_A : \mu_1 \neq \mu_2 \end{array}\right] \text{Testing to compare means}$$

$$\left.\begin{array}{l} \text{Null } H_0 : \sigma_1 = \sigma_2 \\ H_A : \sigma_1 \neq \sigma_2 \end{array}\right] \text{Testing to compare variances.}$$

③ Goodness of Fit Test:

Samples: $X_1 \ldots, X_n$ (iid X). Do they follow some distribution?

- Eg. integer samples $X_i \in \{0, 1, 2 \ldots\}$. Dist. is poisson?
  $H_0 : X \sim Poisson(\lambda)$ H_A : not poisson.
- Cont. samples $X_i \in [-\infty, \infty)$. Dist. is normal?
- ~~Multiple~~ Multinomial $X_i \in \{1, \ldots, M\}$. Dist $\{f_1(\theta), \ldots, f_m(\theta)\}$?

Q.

	Female	Male	Total
Hired	6	12	18
Not Hired	9	25	34
Total	15	37	52

Is there gender bias in hiring?

- Pick 18/52 uniformly at random, $T = M - F$
- Null H_0: Dist. of T as given above
  H_A: Any other dist.

Is 6 a reasonable value for T? How to quantify this?

How to quantify confidence of the testing? P value.

★ The general methodology

$X_1, \ldots, X_n \sim \text{iid } X$

- Test statistic, denoted T. Usually a func. of samples, like sample mean $\overline{X}$, sample variance S^2 etc.
- Acceptance, Rejection regions specified thr. T.
  Eg. Reject H_0 is $T > c$ (right)
  is $T < -c$ (left)
  is $|T| > c$ (two sided)
- α depends on c and distribution for $T|H_0$.
  - Right sided: $\alpha = P(T > c \mid H_0)$
  - Fix α, find c.

DATE

Test for mean (normal samples, known variance)

$X_1, \ldots, X_n \sim \text{iid } N(\mu, 4^2)$

- Test statistic $T = \overline{X} = \frac{1}{n}(X_1 + \ldots + X_n)$
- Null $H_0 : \mu = 0$, $H_A : \mu > 0$.
- Test: Reject H_0 if $T > c$.

SETUP OF A Z TEST

$n = 10$ samplings.

$T = -0.14, -0.25, -1.25, 0.23, 2.96, 3.10, 6.20$

Higher values of T give us more confidence in rejecting null.

$X_1, \ldots, X_{10} \sim \text{iid } N(\mu, 4^2)$

$\alpha = P(\bar{X} > c \mid \mu = 0) \sim N(0, 4^2/10) :\rightarrow$

$$\alpha = P\left(\frac{\overline{X}}{4/\sqrt{10}} > \frac{c}{4/\sqrt{10}}\right) = 1 - F_Z(\sqrt{10}\,c/4)$$

critical value

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001

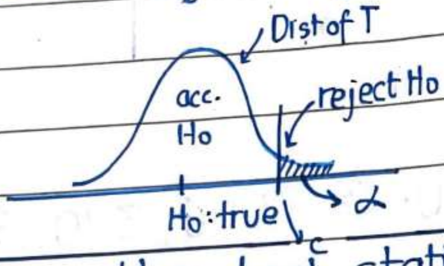
dependent on $n = 10$.

very common

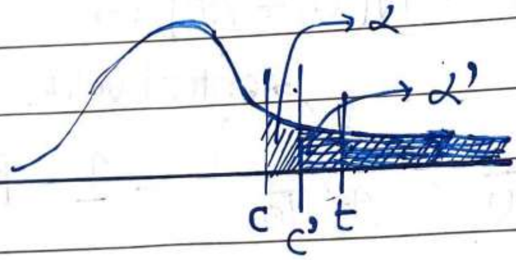
z test at significance level α : Reject H_0 if $T > c$.

★ P Value

c	0	1.62	2.08	2.94	3.26	3.91
α	0.5	0.1	0.05	0.01	0.005	0.001
T = 0.23	Rej	Acc	Acc	Acc	Acc	Acc
T = 2.96	Rej	Rej	Rej	Rej	Acc	Acc
T = 6.20	Rej	Rej	Rej	Rej	Rej	Rej



Suppose the test statistic $T = t$ in one sampling. The lowest significance level α at which the null will be rejected for $T = t$ is said to be the P value of the sampling.



$t > c \Rightarrow$ reject H_0 at α

$t > c' \Rightarrow$ " " " $\alpha' < \alpha$

Crit. value $= t \Rightarrow$ lowest α at which we reject H_0.

Finding P value for $T = t$, put $c = t$ in comp. of α.

T	−0.14	0.23	2.96	6.20
P value	0.544	0.428	0.00964	4.755e−7

If P value is low enough, reject H_0.

Low P value signifies that the sample you have obtained is very unlikely to have come from the distribution under scrutiny.

Q. $X \sim N(\mu, 9)$. For $n=16$ iid X samples, observed sample mean is 10.2. What conclusion would a z test reach if null hypothesis assumes $\mu = 9.5$ (against alternate hypothesis of $\mu > 9.5$) at an α (sig. level) $= 0.05$? What if H_0 assumes $\mu = 8.5$ (against an H_A $\mu > 8.5$?)

$H_0: \mu = 9.5$ $\quad H_A: \mu > 9.5$, $\bar{X} \sim N(\mu, 9/16)$, $\bar{X} = 10.2$

Test: Reject H_0 if $\bar{X} > c$. $\quad \alpha = P(\bar{X} > c \mid \mu = 9.5)$

Now $\dfrac{\bar{X} - 9.5}{(3/4)} \sim Z$ $\qquad = 1 - F_Z\left(\dfrac{c - 9.5}{(3/4)}\right)$

$\Rightarrow c = 9.5 + \frac{3}{4} F_Z^{-1}(0.95) = \underline{10.73}$

Z test @ $\alpha = 0.05$: Accept H_0. To find p value, instead of c, put 10.2. Getting 0.175, which is not low enough to reject H_0.

For $H_0: \mu = 8.5$ $\quad H_A: \mu > 8.5$. $\quad c = 8.5 + \frac{3}{4} F_Z^{-1}(0.95) = \underline{9.73}$

Z test @ $\alpha = 0.05$; → ~~Accept~~ Reject H_0. P value of 0.011.

Q. A face testing app is desired to make accurate identification of faces more than 90% of times in long run. For random 500 photo sample, app makes correct ident. 462 times (92.4% success). What does a z test say abt a null hypothesis that app is only 90% accurate (compared to H_A: app is >90% acc. with $\alpha = 0.05$)

Samples: $X_1, \ldots, X_{500} \sim$ iid Bernoulli(p) $\qquad X_i = \begin{cases} 1; & \text{if app identifies correctly} \\ 0; & \text{otherwise.} \end{cases}$

$\bar{X} = 0.924$

$H_0: p = 0.9$ $\quad H_A: p > 0.9$ $\quad$ Test: Reject H_0 if $\bar{X} > c$

$\bar{X} \sim N(p, p(1-p)/500)$ $\qquad \dfrac{\bar{X} - p}{\sqrt{p(1-p)/500}} \sim Z$

$$\alpha = P(\bar{X} > c \mid p = 0.9) = 1 - F_Z\left(\frac{c - 0.9}{\sqrt{0.09/500}}\right) = 0.05$$

$$c = 0.9 + \sqrt{\frac{0.09}{500}}\, F_Z^{-1}(0.95) = 0.922$$

Since $\bar{X} > c$, z test @ $\alpha = 0.05$: Reject H_0.

p value of 0.036.

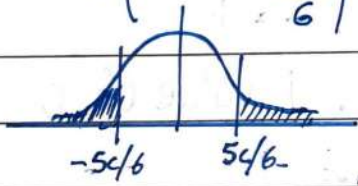
Q. Suppose $X \sim N(\mu, 36)$. For $n = 25$ iid samples of X, observed $\bar{X} = 6.2$. What conclusion would a z test reach if null hypothesis assumes $\mu = 4$, (against $H_A : \mu \neq 4$) at an $\alpha = 0.05$? What if $H_0 : \mu = 8$ (against $H_A : \mu < 8$)?

$H_0 : \mu = 4$, $H_A : \mu \neq 4$, Test: Reject H_0 if $|\mu - 4| > c$

$\bar{X} \sim N(\mu, 36/25)$ $\quad \alpha = P(|\bar{X} - 4| > c \mid \mu = 4) = P\left(\left|\frac{\bar{X} - 4}{6/5}\right| > \frac{c}{6/5}\right)$

$$= P\left(|Z| > \frac{c}{6/5}\right)$$

$$\Rightarrow P\left(|Z| > \frac{5c}{6}\right) = 2F_Z\left(-\frac{5c}{6}\right)$$



$$c = -\frac{6}{5} F_Z^{-1}\left(\frac{0.05}{2}\right) = \underline{\underline{2.352}}$$

$\because |6.2 - 4| = 2.2 < 2.352$, accept H_0.

$H_0 : \mu = 8$, $H_A : \mu < 8$. Test: Reject H_0 if $\bar{X} < c$

$$\alpha = P(\bar{X} < c \mid \mu = 8) = F_Z\left(\frac{c - 8}{6/5}\right) = 0.05$$

$c = 6.02$. $\because \bar{X} = 6.2 > 6.02$, z test @ $\alpha = 0.05$: accept H_0.

DATE

Q. Vaccine hesitancy in a town is reported as 20%. To test this, you call random group of 10 ppl, and see that 3 are hesitant. What is null hypothesis? Will you accept/reject null ~~that~~ at $\alpha = 0.05$ against H_A that fraction > 20%? Power against alternative that fraction is 30%?

Approach 1: Binomial.

Samples: $X_1, \dots, X_{10} \sim \text{iid } Ber(p)$ $\quad X_i = \begin{cases} 1; & \text{if hesitant} \\ 0; & \text{else.} \end{cases}$

$T = X_1 + \dots + X_{10}$.

$H_0: p = 0.2 \quad H_A: p > 0.2 \quad$ Reject H_0 if $T > c$.

$T \sim Binomial(10, p) \quad \alpha = P(T > c \mid p = 0.2)$

$$0.05 = 1 - \sum_{k=0}^{c} \binom{10}{k} p^k (1-p)^{10-k} \Rightarrow c = 4$$

c	α
0	0.893
1	0.624
2	0.322
3	0.121
4	0.033
5	0.0064
6	0.00086

$\because 3 < 4$, accept H_0.

$$\beta = P(\text{Accept } H_0 \mid p = 0.3) = P(T \leq c \mid p = 0.3) = \underline{\underline{0.85}}$$

Approach 2: Normal (100 ppl, 28 hesitant)

~~Samples:~~ $T = Binomial(100, p) \approx Normal(100p, 100p(1-p))$

$$\alpha = P(T > c \mid p = 0.2) = 1 - F_Z\left(\frac{c - 100 \times 0.2}{\sqrt{100 \times 0.2 \times 0.8}}\right) = 0.05.$$

$$\therefore \boxed{c = 26.58}$$

Since $T = 28, > c = 26.58$,

z test @ $\alpha = 0.05$: Reject H_0.

$$\beta = P(T \leq c \mid p = 0.3) = F_Z\left(\frac{26.58 - 100 \times 0.3}{\sqrt{100 \times 0.3 \times 0.7}}\right) = \underline{\underline{0.23}}$$

DATE

Q. $I_{cap.} = 3A$. Because of changes, you suspect $I_{cap} < 3$. Decide to test by measuring I_{cap} of 10 resistors, $\sigma = 0.05A$. If sample mean of measurements $T_{10} < 2.95$, conclude that process is faulty.

① Null, alt. hypothesis? Samples?
② α? ③ If I_{cap} falls to $2.9A$, there could be safety issues. Against H_A of $2.9A$, power $1-\beta$ of test?

Samples: $X_1, \dots, X_{10} \sim N(\mu, 0.05^2)$
$H_0: \mu = 3 \quad H_A: \mu < 3 \quad$ Reject H_0 if $\overline{X} < 2.95$

$$\overline{X} \sim N\left(\mu, \frac{0.05^2}{10}\right) \qquad \alpha = P(\overline{X} < 2.95 \mid \mu = 3)$$

$$= F_Z\left(\frac{2.95-3}{0.05/\sqrt{10}}\right) = 0.00078\ldots$$

$$\beta = P(\overline{X} \geq 2.95 \mid \mu = 2.9) = 1 - F_Z\left(\frac{2.95-2.9}{0.05/\sqrt{10}}\right) = 0.00078\ldots$$

So Power $1-\beta = \boxed{0.99921}$ ✓✓

Refer slides for a crack n resistor version of prob.

Q. Suppose you test n resistors. If sample mean $T_n < c$, you conclude that manufacturing process is faulty. Determine suitable n and c s.t :→
① Significance Level, $\alpha \leq 10^{-6}$
② $\beta \leq 10^{-12}$ (against $H_A: 2.9A$)

$\bar{X} \sim N\left(\mu, \frac{0.05^2}{n}\right)$

$\alpha = P(\bar{X} < c \mid \mu = 3) = F_Z\left(\frac{c-3}{0.05/\sqrt{n}}\right) \leq 10^{-6}$

$\frac{c-3}{0.05/\sqrt{n}} \leq F_Z^{-1}(10^{-6})$ $\quad c \leq 3 + \frac{0.05}{\sqrt{n}}(-4.753)$ — ①

$\beta = P(\bar{X} \geq c \mid \mu = 2.9) = 1 - F_Z\left(\frac{c-2.9}{0.05/\sqrt{n}}\right) \leq 10^{-12}$

$c \geq 2.9 + \frac{0.05}{\sqrt{n}}(7.034)$ — ②

From ① and ②, $n = 35$

Q. Avg. CGPA of students in a college is reported to be 8.0 with standard dev 1. You suspect that average may be 7.5. You decide to sample students to find their CGPA. What sample size do you need for a test at a significance level of 0.05, power 0.95? How will sample size change if you suspect avg as 7.0?

$X_1, \ldots, X_n \sim N(\mu, 1)$. $H_0: \mu = 8$, $H_A: \mu < 8$.

$\bar{X} \sim N\left(\mu, \frac{1}{n}\right)$ Test: Reject H_0 if $\bar{X} < c$

$\alpha = P(\bar{X} < c \mid \mu = 8) = F_Z\left(\frac{c-8}{1/\sqrt{n}}\right) = 0.05$

$c = 8 - \frac{1.645}{\sqrt{n}}$ — ①

$\beta = P(\bar{X} \geq c \mid \mu = 7.5) = 1 - F_Z\left(\frac{c-7.5}{1/\sqrt{n}}\right) = 0.05$

$c = 7.5 + \frac{1.645}{\sqrt{n}}$ — ② $\Rightarrow n \approx 44$

 If 7, $c = \cancel{7} + \frac{1.645}{\sqrt{n}} \Rightarrow n \approx 11$.