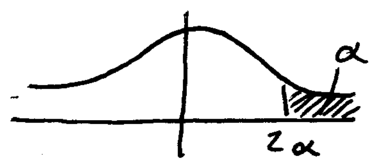
MARCH 19/18
APPLIED ANAL

Last time - Continuous random variable
- Probability density function $\int_c^d f(x)dx = Pr(x \in [c,d])$
- $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x) = 1$
- mean, variance, standard deviation
- distribution function
- normal distribution
- standard normal

Let $0 < \alpha < 1$, then Z_α is a number such that
$Pr(Z > Z_\alpha) = \alpha$



$Pr(Z \leq Z_\alpha) = 1 - \alpha$, To find Z_α we look inside the table for a number near $1-\alpha$

- e.g. Find (i) $Z_{.01}$ (ii) $Z_{.05}$ (iii) $Z_{.83}$
(i) $F(2.32) = .9888$, $F(2.33) = .9901$, $Z_{.01} = 2.325$
(ii) $F(1.64) = .9495$, $F(1.65) = .9505$, $Z_{.05} = 1.645$
(iii) $F(-0.95) = .1711$, $F(-0.96) = .1685$, $Z_{.83} = -0.955$

Suppose x is normal with new μ, standard deviation σ
Then we can <u>standardize</u> : $Z = \frac{x-\mu}{\sigma}$ is standard normal
$Pr(a \leq X \leq b) = Pr\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$
$= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$

- e.g Let x be normal with mean 70, standard dev. 5 :
Find (i) $Pr(x \leq 80)$ (ii) $Pr(x \leq 66)$ (iii) $Pr(69 \leq x \leq 72)$
(i) $F\left(\frac{80-70}{5}\right) = F(2) = .9972$
(ii) $1 - Pr(x \leq 66) = 1 - F\left(\frac{66-70}{2}\right) = 1 - F(-0.8) = 1 - 0.2119 = 0.7881$
(iii) $F\left(\frac{72-70}{2}\right) - F\left(\frac{69-70}{2}\right) = F(.4) - F(.2) = .6554 - .4201 = .2347$

- e.g. Let x be normal, $\mu = 100$, $\sigma = 10$, Find
(i) $P_r(x < 80)$ (ii) a so that $P_r(x > a) = .3$
(i) $F\left(\frac{80-100}{10}\right) = F(-2) = 0.0228$
(ii) $P_r(X \leq a) = .7$, so $F\left(\frac{a-100}{10}\right) = 0.7$
$F(0.525) = 0.7$, so $\frac{a-100}{10} = 0.525$, so $a = 105.25$

- e.g. We have a machine that fills jars with jelly beans. Our machine can put in amounts that we normally distributed with $\sigma = 3g$
How should we set μ so that only 2% of the jars contain less than 500g of jelly beans?

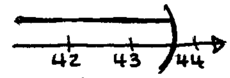
$P_r(x < 500) = .02$
$P_r\left(z < \frac{500-\mu}{3}\right) = .02$
$F(-2.055) = .02$, so $\frac{500-\mu}{3} = -2.055$, $\mu = 506$

Let x be a binomial, then
$$\frac{x - np}{\sqrt{np(1-p)}} \longrightarrow Z \text{ as } n \to \infty$$
If $np \geq 15$ and $n(1-p) \geq 15$, then it is reasonable to use a normal aprox. to a binomial
We must adjust our endpoints ± 0.5 to include or exclude the endpoint.

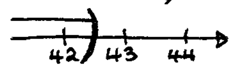
- e.g. Flip a balanced coin 100 times, count the heads.
Find the prob of getting:
(i) at most 43 heads
(ii) fewer than 43 heads
(iii) at least 43 heads
(iv) more than 43 heads
(v) at least 38, no more than 43 heads
(vi) exactly 43 heads (using normal approx.)

$$Z = \frac{x - np}{\sqrt{np(1-p)}} = \frac{x - 100(1/2)}{\sqrt{100(1/2)(1/2)}} = \frac{x-50}{5}$$

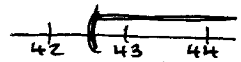
(i) $F\left(\frac{43.5-50}{5}\right) = F(-1.3) = .0968$



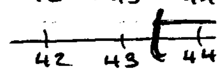
(ii) $F\left(\frac{42.5-50}{5}\right) = F(-1.5) = .0668$



(iii) $1 - F\left(\frac{42.5-50}{5}\right) = 1 - .0668 = .9332$

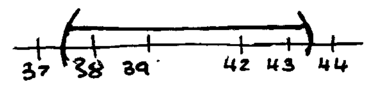


(iv) $1 - F\left(\frac{43.5-50}{5}\right) = 1 - .0968 = .9032$



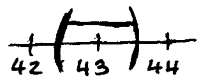
(v) $F\left(\frac{43.5-50}{5}\right) - F\left(\frac{37.5-50}{5}\right)$

$\Rightarrow 0.0968 - F(-2.5) = .0968 - .0062 = .0906$



(vi) $F\left(\frac{43.5-50}{5}\right) - F\left(\frac{42.5-50}{5}\right)$

$\Rightarrow .0968 - .0668 = .03$



- e.g. Roll a balanced die 720 times, Find the prob. of getting exactly 110 fours

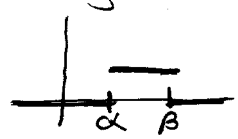
$$Z = \frac{x - np}{\sqrt{np(1-p)}} = \frac{x - 720(1/6)}{\sqrt{720(1/6)(5/6)}} = \frac{x-120}{10}$$

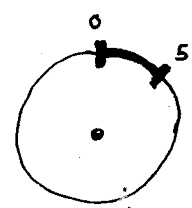
$$F\left(\frac{110.5-120}{10}\right) - F\left(\frac{109.5-120}{10}\right)$$

$\Rightarrow F(-0.95) - F(-1.05) = .1711 - .16109 = .0242$

The <u>uniform distribution</u> has density function

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & x \in [\alpha, \beta] \\ 0, & \text{elsewhere} \end{cases}$$





- e.g. We take a wheel with radius 10 cm. We have painted a 5cm strip around the circumferance orange. We spin the wheel radially on its axis. Blindfolded, I throw a dart and (...?) the circumference of the wheel. Find the prob. I hit the orange strip.

$$f(x) = \begin{cases} \frac{1}{10\pi}, & x \in (0, 10\pi) \\ 0, & \text{elsewhere} \end{cases}$$

$$Pr(0 \le x \le 5) = \int_0^5 \frac{1}{10\pi}\,dx = \frac{5}{10\pi}$$

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx = \int_\alpha^\beta \frac{x}{\beta-\alpha}\,dx = \left.\frac{x^2}{2(\beta-\alpha)}\right|_\alpha^\beta = \frac{\beta^2-\alpha^2}{2(\beta-\alpha)}$$

$$= \frac{\cancel{(\beta-\alpha)}(\beta+\alpha)}{(2)\cancel{(\beta-\alpha)}} = \frac{\beta+\alpha}{2}$$

$$\sigma^2 = \frac{(\beta-\alpha)^2}{12}$$

MARCH 21/18
APPLIED ANAL.

- last time
  - Z_α $\Pr(Z > Z_\alpha) = \alpha$
  - standardizing $Z = \frac{x-\mu}{\sigma}$
  - normal approx. to
    $Z = \frac{x-np}{\sqrt{np(1-p)}}$, adjust endpoints $\pm \frac{1}{2}$
  - uniform distribution

Let $\alpha \in \mathbb{R}$, $\beta > 0$, we say that x has log-normal dist.
if its density function is:

$$f(b) = \begin{cases} \frac{1}{\sqrt{2\pi}\beta}\left(x^{-1}e^{-((\ln x)-\alpha)^2/(2\beta^2)}\right), & x > 0 \\ 0, & x \le 0 \end{cases}$$

$\ln x$ is normal with mean α, standard deviation β

$$\Pr(a \le x \le b) = F\left(\frac{\ln b - \alpha}{\beta}\right) - F\left(\frac{\ln a - \alpha}{\beta}\right)$$

- e.g. x has log-normal distribution, $\alpha = 5$, $\beta = 2$
  Find $\Pr(e^4 \le x \le e^7)$

$$F\left(\frac{\ln(e^7) - 5}{2}\right) - F\left(\frac{\ln(e^4) - 5}{2}\right) = F\left(\frac{7-5}{2}\right) - F\left(\frac{4-5}{2}\right)$$
$$= F(1) - F(-0.5) = 0.8413 - 0.3085 = 0.5328$$
$$\mu = e^{\alpha + (\beta/2)}, \quad \sigma^2 = (e^{2\alpha + \beta^2})(e^{(\beta^*}-1)$$

Chapter 7

A population is a collection of numbers

The population is normal if upon selecting a number at random and letting it be x, the variable x is a normal variable.

The population mean μ is the mean of x

The population variance σ^2 is the variance of x

The population standard deviation is $\sigma = \sqrt{\sigma^2}$

A Parameter is a number determined from a population (e.g. μ, σ^2)

A statistic is a number calculated from a random sample. (e.g. $\bar{x}$, s^2)

A statistic is called an unbiased estimator if the expected value of the statistic equals the parameter

- e.g. $\bar{x}$ is an unbiased estimator to μ

We want the estimator to be as efficient as possible, that is, to have a σ as small as possible.

$\bar{x}$ is an efficient estimator for μ.

If we have a sample of size n, the expected value of $\bar{x}$ is μ.

The standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$

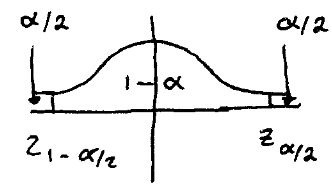
we concen ourseives with μ, $\bar{x}$

$\bar{x}$ is our point estimate for μ

If x is normal, so is $\bar{x}$, and so

$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ ... is standard normal.

For any α, $0 < \alpha < 1$



$Pr[-z_{\alpha/2} \le Z \le z_{\alpha/2}] = 1-\alpha$

$Pr[|Z| \le z_{\alpha/2}] = 1-\alpha$

$Pr\left[\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| < z_{\alpha/2}\right] = 1-\alpha$

$Pr\left[|\bar{x}-\mu| \le \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right] = 1-\alpha$

Let $E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$ we call this the maximum error of the estimate

(with probability $1-\alpha$)

- e.g. We have a normal population with standard dev. zero.

We take a radom sample of size 100. Find the max error of the estimate for μ with probability .99

$1-\alpha = .99$, so $\alpha = .01$, $Z_{.005} = 2.575$

$E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} = \frac{2.575(20)}{\sqrt{100}} = 5.15$

Suppose E is predetermined, and we want to find n

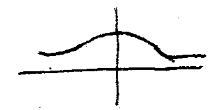
$E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$, so $\sqrt{n} = \frac{z_{\alpha/2}\sigma}{E}$, and $n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$

- e.g. Suppose we have a normal population with $\sigma = 10$. How large a sample do we need to get a max. error of the estimate of 4, with prob. 0.99.

$n = \left(\frac{2.575\,(10)}{4}\right)^2$, 41.4 round up to 42

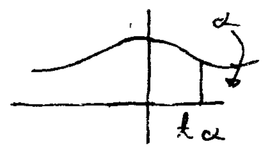
If σ is unknown, we want to use s to approximate σ

Let $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ - This has t-distribution with $\nu = n-1$ degrees of freedom.



t is symmetric, non-zero, Standard deviation is not 1

$t \to Z$ as $n \to \infty$



The table gives t_α values

$Pr(t > t_\alpha) = \alpha$

$t_{1-\alpha} = -t_\alpha$

We will use $E = \frac{t_{\alpha/2}\, s}{\sqrt{n}}$

- e.g. We have a normal population. We take a random sample of size 25. Find the max. error of the estimate with prob. .95, assuming that the standard dev. is 12.

$1 - \alpha = .95$, so $\alpha = .05$, $t_{.025} = 2.064$ ($\nu = 24$)

$$E = \frac{2.064\,(12)}{\sqrt{25}}$$

→ If $n \geq 30$ we use Z even if σ is unknown

→ If σ is known, $E = \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}$

→ If σ is unknown, $n < 30$, $E = \frac{t_{\alpha/2}\, s}{\sqrt{n}}$, $\nu = n-1$

→ If σ is unknown, $n \geq 30$, $E = \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}$

- e.g. We take a random sample of size 100 from a normal pop., we get a sample standard deviation of 7. Find the max error of the estimate with prob. 0.96

→ σ is unknown, $n \geq 30$, so $E = \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}}$

$1 - \alpha = 0.95$, $\alpha = 0.05$, $z_{.025} = 1.96$

$$E = \frac{1.96\,(7)}{\sqrt{100}}$$

→ x has density function

$$f(x) = \begin{cases} K\sin x, & x \in [0, \pi] \\ 0, & \text{elsewhere} \end{cases}$$

(i) Find K

(ii) Find $Pr(x > \pi/4)$

↪ (i) $1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^{\pi} K\sin x = -K\cos x \Big|_0^{\pi} = -K(-1-1)$, $K = 1/2$

(ii) $\int_{\pi/4}^{\infty} f(x)dx = \int_{\pi/4}^{\pi} 1/2 \sin x dx = -1/2 \cos x \int_{\pi/4}^{\pi} = -1/2(-1 - 1/\sqrt{2})$

→ Find $Z_{.03}$

$Pr(Z > Z_{.03}) = .03$

$Pr(Z < Z_{.97}) = .97$

$F(1.88) = .9699$, $F(1.89) = .9706$

$Z_{.03} = 1.885$

→ x is normal, $\mu = 80$, $\sigma = 10$

Find (i) $Pr(x < 63)$

(ii) a so that $Pr(x < a) = 0.3$

↪ (i) $F\left(\frac{63-80}{10}\right) = F(-1.7) = 0.0446$ %

(ii) $F\left(\frac{a-80}{10}\right) = 0.3$

$F(-0.525) = 0.7$

$\frac{a-80}{10} = -0.525$, $a = 80 - 5.25$

→ x is normal, $\mu = 100$

$Pr(x > 120) = 0.1$

Find σ

$Pr(x < 120) = 0.9$

$F\left(\frac{120-100}{\sigma}\right) = 0.9$ → $F(1.285) = 0.9$

$20/\sigma = 1.285 \Rightarrow \sigma = 20/1.285$

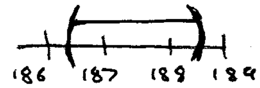
→ We flip a balanced coin 400 times. Find the prob. of getting either 187 or 188 heads

(i) exactly

(ii) a normal approx.

↳ (i) $\binom{400}{187}(1/2)^{187}(1/2)^{213} + \binom{400}{188}(1/2)^{188}(1/2)^{212}$

(ii) $Z = \frac{x-np}{\sqrt{np(1-p)}} = \frac{x-400(1/2)}{\sqrt{400(1/2)(1/2)}} = \frac{x-200}{10}$



$F\left(\frac{188.5-200}{10}\right) - F\left(\frac{186.5-200}{10}\right) = F(-1.15) - F(-1.35)$

$= .1251 - .0885$

→ We take a random sample of size n from a normal population. What is the max. error of the estimate for μ if:

(i) $\sigma = 10$, $n = 16$

(ii) σ is unknown, $s = 12$, $n = 16$

(iii) σ is unknown, $s = 12$, $u = 1000$ → "n"

↳ in each case with prob. .95

(i) as σ is known, $E = \frac{z_{\alpha/2}\,\sigma}{\sqrt{n}} = \frac{z_{.025}(10)}{\sqrt{16}} = \frac{1.96(10)}{4}$

(ii) as σ is unknown, $n < 30$, $E = \frac{t_{\alpha/2}\,s}{\sqrt{n}} = \frac{t_{.025}(12)}{\sqrt{16}} = \frac{2.131(12)}{4}$

(iii) as σ is unknown, $n \geq 30$, $E = \frac{z_{\alpha/2}\,s}{\sqrt{n}} = \frac{z_{..}(12)}{\sqrt{1000}} = \frac{1.96(12)}{\sqrt{1000}}$

- e.g. When a normal pop $\sigma = 30$, how large a sample do we need to get a max. error in the estimate of μ to be 2 with prob. 0.95?

$n = \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2 = \left(\frac{1.96(30)}{2}\right)^2 = 864.36$ so 865

MARCH 23 /18
APPLIED ANAL.

Last time
- log-normal distribution
- population
- parameter, Statistic
- max error of the estimate
  - σ known, $E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$
  - σ unknown, $n < 30$, $E = \frac{t_{\alpha/2}\sigma}{\sqrt{n}}$; $\nu = n-1$
  - σ unknown, $n \geq 30$, $E = \frac{z_{\alpha/2} S}{\sqrt{n}}$
- if we have σ, and we want a particular n, $n = \left(\frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right)^2$

we have seen that $Pr(|\bar{x} - \mu| \leq E) = 1 - \alpha$

The prob. that $\bar{x}$ is between $\mu - E$ and $\mu + E$ is $1 - \alpha$

we would like to say that the prob. that μ is between $\bar{x} - E$ and $\bar{x} + E$ is $1 - \alpha$, but μ is either there or it isn't!

Instead we will say that we are $100(1-\alpha)\%$ confident that μ is between $\bar{x} - E$ and $\bar{x} + E$

We call $(\bar{x} - E, \bar{x} + E)$ the $100(1-\alpha)\%$ confidence interval.

- e.g we take a random sample of size 12 from a normal pop. and get a sample mean of 40 and a sample standard deviation of 7. Find the 99% confidence interval for μ.

As σ is unknown, $n < 30$, $E = \frac{t_{\alpha/2} S}{\sqrt{n}} = \frac{t_{.005}(7)}{\sqrt{12}} = \frac{3.106(7)}{\sqrt{12}}$ $(\nu = 11)$

$[\bar{x} - \sqrt{2}, \bar{x} + \sqrt{2}] = \left[40 - \frac{3.106(7)}{\sqrt{2}}, 40 + \frac{3.106(7)}{\sqrt{2}}\right]$

- e.g. Same problem, except that we know $\sigma = 7$

As σ is known, $\sqrt{2} = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} = \frac{2.005(7)}{\sqrt{12}} = \frac{2.575(7)}{\sqrt{12}}$

$[\bar{x} - \sqrt{2}, \bar{x} + \sqrt{2}] = \left[40 - \frac{2.575(7)}{\sqrt{2}}, 40 + \frac{2.575(7)}{\sqrt{2}}\right]$

We want to make hypothesis concerning means

The <u>null hypothesis</u>, H_0, is the hypothesis that we are setting up to see if we can reject it.

The alternative hypothesis, H_1, is the logical negative of H_0

It is what we are trying to establish.

We will either reject H_0 or fail to reject H_0

Never accept H_0!

A type I error occurs if we reject H_0, even though it is true

A type II error occurs if we had to reject H_0, even though it is false

	H_0 true	H_0 false
reject H_0	Type I error	✓
do not reject H_0	✓	Type II error

Possible hypothesis:

$H_0 : \mu \le \mu_0$	$H_0 : \mu \ge \mu_0$	$H_0 : \mu = \mu_0$
$H_1 : \mu > \mu_0$	$H_1 : \mu < \mu_0$	$H_1 : \mu \ne \mu_0$

The possibility of a type I error (assuming $\mu = \mu_0$) is called the level of significanse ...

The probability of a type II error, for some ... is denoted β.

In most cases, α will be given and we design the experiment accordingly. But we can have:

- e.g. Suppose we have a normal population with $\sigma = 10$

We wish to test this: $H_0 : \mu \ge 50$, $H_1 : \mu \le 50$

We have a sample of size 25 and reject H_0 if $\bar{x} < 47$. Find:

(i) Find α (ii) Find β if $\mu = 45$

↳ (i) Assuming $\mu = 50$, we reject (incorrectly) is $\bar{x} < 47$

$$\Pr[\bar{x} < 47] = \Pr\left[\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{47-50}{10/\sqrt{25}}\right] = \Pr(Z < -1.5) = 0.0668$$

(ii) $\Pr(\bar{x} > 47) = 1 - \Pr(\bar{x} < 47) = 1 - \Pr\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le \frac{47-45}{10/\sqrt{25}}\right)$

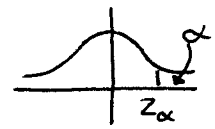
$\Rightarrow 1 - \Pr(Z \le 1) = 1 - 0.8413 = 0.1587$

Suppose we are given α, and for now, that σ is known

$H_0 = \mu \le \mu_0$ | Let $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$; we will reject that if

$H_1 = \mu > \mu_0$

we reject if $Z > Z_\alpha$



- e.g. the number of marshmallows in boxes of Krusty O's is normally distributed with $\sigma = 10$. Krusty claims that the number of marshmallows per box is on average at most 60. We randomly select 100 boxes and put a sample mean of 63. Test the claim at a .05 level of significance

$H_0 : \mu \le 60$ As σ is known, let $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ and

$H_1 : \mu > 60$ reject if $Z > Z_{.05}$

$Z = \frac{63-60}{10/\sqrt{100}} = 3$; $Z_{.05} = 1.645$

As $Z > Z_{.05}$ we reject the claim of a .05 level of significance.

The book likes to do things like : $H_0 : \mu = 50$ ← $H_0 : \mu \le 50$

$H_1 : \mu \ge 50$ ← correct