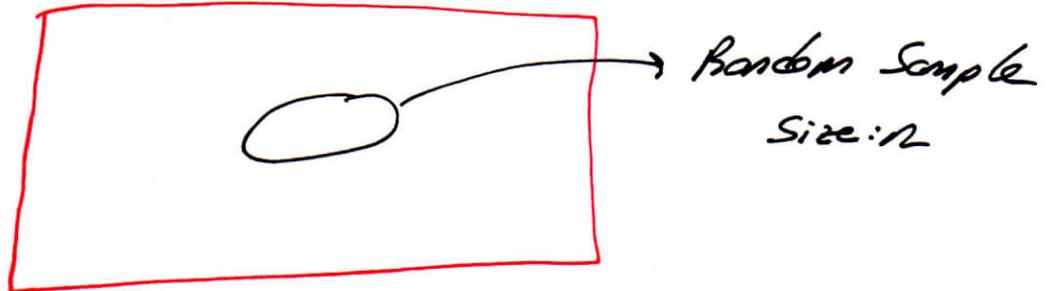




ENGINEERS STAT LECTURE NOTES	CHAPTER 6
---------------------------------	-----------

* Basic Concepts:



Population $N = 12000$

X_i : Weekly food expenditure of a student at Bilkent University

$Y_i = \begin{cases} 1 & \text{if the student makes any sport} \\ 0 & \text{o.w.} \end{cases}$

Population Parameters

(Unknown Constants)

Sample Statistics

(Known Variables)

MEAN

μ

$$\bar{X} = \frac{\sum X_i}{n}$$

VARIANCE

σ^2

$$S^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n-1}$$

standard deviation

σ

$$s = \sqrt{S^2}$$

PROPORTION

p

$$\hat{p} = \frac{\sum Y_i}{n} = \frac{T}{n}$$

$T \sim \text{Binomial}(n, p)$

①



SAMPLING DISTRIBUTIONS

Sample Mean: \bar{X}

* If X_i has a normal distribution with mean μ and variance σ^2 , we know (from prob.) that linear combinations of Normal Random Variables have also Normal Distribution. Then;

$$X_i \stackrel{i.i.d}{\sim} \text{Normal}(\mu; \sigma^2) \Rightarrow \bar{X} \sim \text{Normal}\left(\mu; \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$

Then, sampling distribution of \bar{X} is Normal with mean μ and variance $\frac{\sigma^2}{n}$. It is easy to show that $E(\bar{X}) = \mu$ & $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

(Remember $E(aX + bY + c) = aE(X) + bE(Y) + c$ and $\text{Var}(aX + bY + c) = \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$ for X, Y independent Random Variables and a, b, c constants)

* If X_i has ANY distribution with mean μ and variance σ^2 , then for large n (practically $n > 30$) \bar{X} has again Normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. This is called "Central Limit Theorem"

$$X_i \stackrel{i.i.d}{\sim} \text{ANY DISTRIBUTION (Large } n) \Rightarrow \bar{X} \sim \text{Normal}\left(\mu; \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$



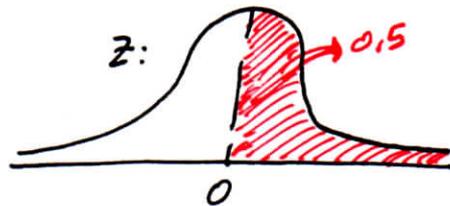
* Remember, we find Normal Probabilities from Z : Standard Normal distribution table.

For a single unit, we have $Z = \frac{X - \mu}{\sigma}$

Then, for sample mean, we have $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Note that $Z \sim \text{Normal}(\mu=0; \sigma^2=1^2)$

* Finding Probabilities from Z -table:

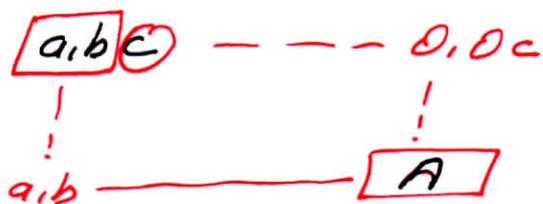
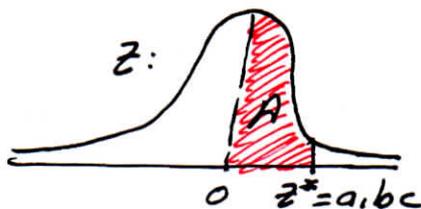


We have; (i) Total Area = 1

(ii) Half Area = 0,5

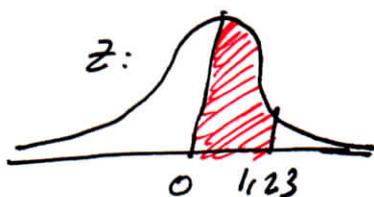
(iii) Symmetric areas are equal.

$$P(-z_2 < Z < z_1) = P(z_1 < Z < z_2)$$

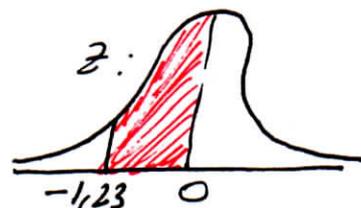


$$A = P(0 < Z < a, b, c)$$

(I) $P(0 < Z < 1,23) = ?$



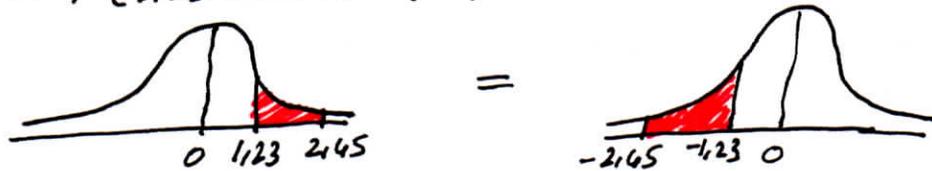
=



$$P(0 < Z < 1,23) = 0,3907 = P(-1,23 < Z < 0)$$

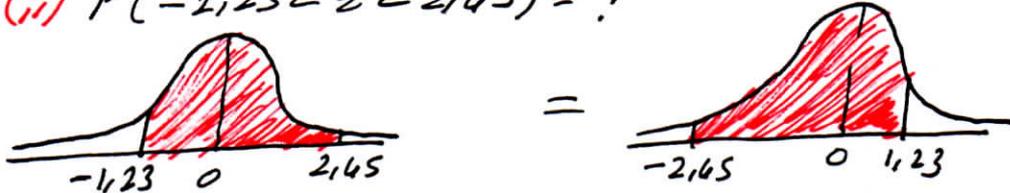


(II) (i) $P(1,23 < z < 2,45) = ?$



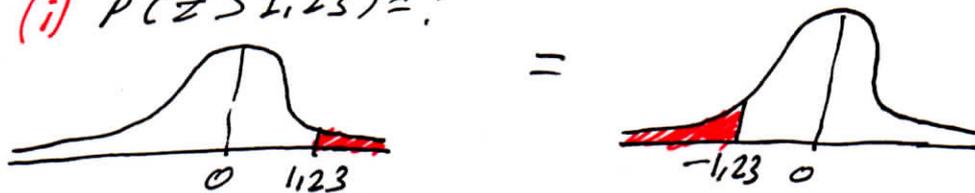
$$P(1,23 < z < 2,45) = 0,4929 - 0,3907 = 0,1022 = P(-2,45 < z < -1,23)$$

(ii) $P(-1,23 < z < 2,45) = ?$



$$P(-1,23 < z < 2,45) = 0,4929 + 0,3907 = 0,8836 = P(-2,45 < z < 1,23)$$

(III) (i) $P(z > 1,23) = ?$



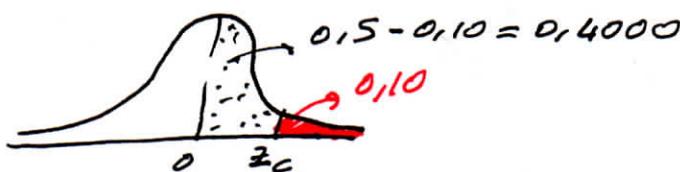
$$P(z > 1,23) = 0,5 - 0,3907 = 0,1093 = P(z < -1,23)$$

(ii) $P(z < 2,45) = ?$



$$P(z < 2,45) = 0,5 + 0,4929 = 0,9929 = P(z > -2,45)$$

IV $P(z > z_c) = 0,10 \Rightarrow z_c = ?$



$$P(0 < z < 1,28) = 0,3997 \Rightarrow z_c = 1,28$$

6.9 Shear-strength measurements for spot welds of a certain type have been found to have a standard deviation of approximately 10 psi. If 100 test welds are to be measured, find the approximate probability that the sample means will be within 1 psi of the true population mean.

6.10 If shear-strength measurements have a standard deviation of 10 psi, how many test welds should be used in the sample if the sample mean is to be within 1 psi of the population mean with probability approximately 0.95?

6.9) $X \sim \text{Normal}(\mu; \sigma^2 = 10^2) \quad n = 100$ (Even if X is NOT normally distributed, $n > 30$ and by CLT, $\bar{X} \sim \text{Normal}$)

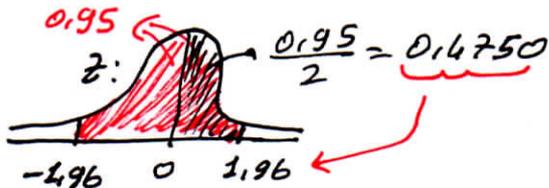
$$P(\bar{X} \in (\mu - 1; \mu + 1)) = P(\mu - 1 < \bar{X} < \mu + 1)$$

$$= P\left(\frac{\mu - 1 - \mu}{\sigma/\sqrt{n}} < \underbrace{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}_{= z} < \frac{\mu + 1 - \mu}{\sigma/\sqrt{n}}\right) = P\left(-\frac{\sqrt{100}}{10} < z < \frac{\sqrt{100}}{10}\right)$$

$$= P(-1 < z < 1) = 2 \cdot 0.3413 = 0.6826$$



6.10) $P(\mu - 1 < \bar{X} < \mu + 1) = \dots = P\left(-\frac{\sqrt{n}}{10} < z < \frac{\sqrt{n}}{10}\right) = 0.95$



Then; $\frac{\sqrt{n}}{10} = 1.96$

$$n = (1.96)^2 = 384.16$$

$n = 385$ ← ROUND UP!

6.13 Resistors of a certain type have resistances that average 200 ohms with a standard deviation of 10 ohms. Twenty-five of these resistors are to be used in a circuit.

- Find the probability that the average resistance of the 25 resistors is between 199 and 202 ohms.
- Find the probability that the total resistance of the 25 resistors does not exceed 5,100 ohms. [Hint: Note that $P(\sum_{i=1}^n X_i > a) = P(n\bar{X} > a) = P(\bar{X} > a/n)$.]
- What assumptions are necessary for the answers in (a) and (b) to be good approximations?

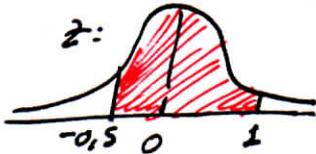
6.13) c) Assume $X \sim \text{Normal}(\mu = 200; \sigma^2 = 10^2)$ ~~and~~

$n = 25$



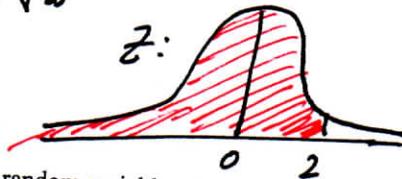
$$a) P(199 < \bar{X} < 202) = P\left(\frac{199-200}{10/\sqrt{25}} < \underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{=z} < \frac{202-200}{10/\sqrt{25}}\right)$$

$$= P(-0,5 < z < 1) = 0,1915 + 0,3413 = 0,5328$$



$$b) P(\sum X_i \leq 5100) = P\left(\frac{\sum X_i}{n} \leq \frac{5100}{25}\right) = P(\bar{X} \leq 204)$$

$$= P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{204-200}{10/\sqrt{25}}\right) = P(z \leq 2) = 0,5 + 0,4772 = 0,9772$$



6.17 The strength of a thread is a random variable with mean 0.5 pound and standard deviation 0.2 pound. Assume the strength of a rope is the sum of the strengths of the threads in the rope.

- Find the probability that a rope consisting of 100 threads will hold 45 pounds.
- How many threads are needed for a rope that will hold 50 pounds with 99% assurance?

6.17) X : strength of a thread
 $\mu = 0,5$; $\sigma = 0,2$;

$$a) P\left(\sum_{i=1}^{100} X_i > 45\right) = P\left(\frac{\sum X_i}{n} > \frac{45}{100}\right) = P(\bar{X} > 0,45)$$

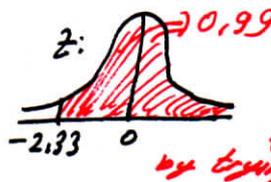
$$= P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{0,45-0,50}{0,2/\sqrt{100}}\right) = P(z > -2,5) = 0,5 + 0,4938 = 0,9938$$



$$b) P\left(\sum_{i=1}^n X_i > 50\right) = P\left(\frac{\sum X_i}{n} > \frac{50}{n}\right) = P\left(\bar{X} > \frac{50}{n}\right) = P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > \frac{50-0,5n}{0,2/\sqrt{n}}\right)$$

$$= P\left(z > \frac{50-0,5n}{0,2\sqrt{n}}\right) = 0,99$$

$\frac{50-0,5n}{0,2\sqrt{n}} = -2,33$
 $50 = 0,5n - 0,466\sqrt{n}$
 $n = 110$





6.18 Many bulk products, such as iron ore, coal, and raw sugar, are sampled for quality by a method that requires many small samples to be taken periodically as the material is moving along a conveyor belt. The small samples are then aggregated and mixed to form one composite sample. Let Y_i denote the volume of the i th small sample from a particular lot, and suppose Y_1, \dots, Y_n constitutes a random sample with each Y_i having mean μ and variance σ^2 . The average volume of the samples μ can be set by adjusting the size of the sampling device. Suppose the variance of sampling volumes σ^2 is known to be approximately 4 for a particular situation (measurements are in cubic inches). It is required that the total volume of the composite sample exceed 200 cubic inches with probability approximately 0.95 when $n = 50$ small samples are selected. Find a setting for μ that will allow the sampling requirements to be satisfied.

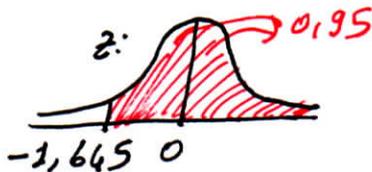
$$6.18) \sigma^2 = \text{Var}(Y_i) = 4 \quad \mu = E(Y_i) \quad n = 50$$

$$P\left(\sum_{i=1}^{50} X_i > 200\right) = 0,95$$

$$P\left(\sum_{i=1}^{50} X_i > 200\right) = P\left(\frac{\sum X_i}{n} > \frac{200}{50}\right) = P(\bar{X} > 4)$$

$$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{4 - \mu}{2/\sqrt{50}}\right) = P\left(z > \frac{4 - \mu}{0,2828}\right) = 0,95$$

$= -1,645$



$$4 - \mu = 0,2828 \cdot (-1,645)$$

$$\mu = 4 + 0,2828 \cdot 1,645$$

$$\boxed{\mu = 4,465}$$

6.21 Suppose that X_1, \dots, X_{n_1} and Y_1, \dots, Y_{n_2} constitute independent random samples from populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 , respectively. Then the Central Limit Theorem can be extended to show that $\bar{X} - \bar{Y}$ is approximately normally distributed for large n_1 and n_2 , with mean $\mu_1 - \mu_2$ and variance $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$.

Water flow through soils depends, among other things, on the porosity (volume proportion due to voids) of the soil. To compare two types of sandy soil, $n_1 = 50$ measurements are to be taken on the porosity of soil A, and $n_2 = 100$ measurements are to be taken on soil B. Assume that $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.02$. Find the approximate probability that the difference between the sample means will be within 0.05 unit of the true difference between the population means, $\mu_1 - \mu_2$.

6.22 Refer to Exercise 6.21. Suppose samples are to be selected with $n_1 = n_2 = n$. Find the value of n that will allow the difference between the sample means to be within 0.04 unit of $\mu_1 - \mu_2$ with probability approximately 0.90.



6.21)

$$\frac{X_i}{n_1} \quad \frac{Y_i}{n_2}$$

$$E(X_i) = \mu_1 \quad E(Y_i) = \mu_2$$

$$\text{Var}(X_i) = \sigma_1^2 \quad \text{Var}(Y_i) = \sigma_2^2$$

$$(\bar{X} - \bar{Y}) \sim \text{Normal} \left(\mu_1 - \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)$$

Soil A

$$n_1 = 50$$

$$\sigma_1^2 = 0,01$$

Soil B

$$n_2 = 100$$

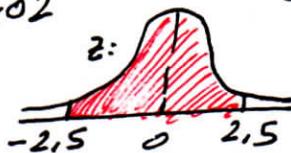
$$\sigma_2^2 = 0,02$$

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$P(\mu_1 - \mu_2 - 0,05 < \bar{X} - \bar{Y} < \mu_1 - \mu_2 + 0,05)$$

$$= P\left(\frac{(\mu_1 - \mu_2 - 0,05) - (\mu_1 - \mu_2)}{\sqrt{\frac{0,01}{50} + \frac{0,02}{100}}} < Z < \frac{(\mu_1 - \mu_2 + 0,05) - (\mu_1 - \mu_2)}{\sqrt{\frac{0,01}{50} + \frac{0,02}{100}}} \right)$$

$$= P\left(-\frac{0,05}{0,02} < Z < \frac{0,05}{0,02} \right) = P(-2,5 < Z < 2,5)$$

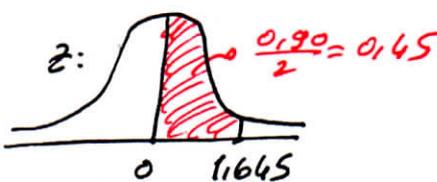


$$= 2 \cdot 0,4938 = 0,9876$$

6.22)

$$P\left(\frac{-0,04}{\sqrt{\frac{0,01}{n} + \frac{0,02}{n}}} < Z < \frac{0,04}{\sqrt{\frac{0,01}{n} + \frac{0,02}{n}}} \right) = 0,90$$

$= 1,645$



$$\sqrt{\frac{0,01}{n} + \frac{0,02}{n}} = \frac{0,04}{1,645}$$

$$\sqrt{n} = \frac{1,645 \cdot \sqrt{0,03}}{0,04} = 7,123$$

$$n = 50,74 \Rightarrow \boxed{n = 51}$$

6.23 An experiment is designed to test whether operator A or operator B gets the job of operating a new machine. Each operator is timed on 50 independent trials involving the performance of a certain task on the machine. If the sample means for the 50 trials differ by more than 1 second, the operator with the smaller mean gets the job. Otherwise, the experiment is considered to end in a tie. If the standard deviations of times for both operators are assumed to be 2 seconds, what is the probability that operator A gets the job even though both operators have equal ability?

6.23)

Operator A

$$n_1 = 50$$

$$\sigma_1 = 2$$

$$\mu_1 = \mu$$

Operator B

$$n_2 = 50$$

$$\sigma_2 = 2$$

$$\mu_2 = \mu$$

$$(\bar{X}_B - \bar{X}_A) \sim \text{Normal} \left(\underbrace{0}_{=\mu - \mu}; \underbrace{\frac{2^2}{50} + \frac{2^2}{50}}_{=\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

$$P(\bar{X}_B > \bar{X}_A + 1) = P(\bar{X}_B - \bar{X}_A > 1)$$

$$= P\left(\frac{(\bar{X}_B - \bar{X}_A) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > \frac{1}{\sqrt{\frac{2^2}{50} + \frac{2^2}{50}}}\right) = P(Z > 2.5) =$$

$$= 0.5 - 0.4938 = 0.0062$$

Sample Proportion: \hat{p}

* Remember; $y_i = \begin{cases} 1 & \text{if student makes sport} \\ 0 & \text{o.w.} \end{cases}$

Then, p : Probability of a randomly selected student to make sport.

$T = \sum_{i=1}^n y_i$: # of success among n trials

$T \sim \text{Binomial}(n, p)$

$\mu = E(T) = n \cdot p$; $\sigma^2 = \text{Var}(T) = n \cdot p \cdot (1-p)$



* The sample proportion is; $\hat{p} = \frac{T}{n}$.

Then; $\mu_{\hat{p}} = E(\hat{p}) = E\left(\frac{T}{n}\right) = \frac{1}{n} E(T) = \frac{1}{n} \cdot np = p$

$$\sigma_{\hat{p}}^2 = \text{Var}(\hat{p}) = \text{Var}\left(\frac{T}{n}\right) = \frac{1}{n^2} \text{Var}(T) = \frac{1}{n^2} \cdot n \cdot p \cdot (1-p)$$

$$= \frac{p \cdot (1-p)}{n}$$

If n is large enough; practically $p \pm 2\sigma_{\hat{p}}$ lies in the interval $(0,1)$, Normal approximation to Binomial distribution is applicable. Namely; Sampling distribution of \hat{p} is;

$$\hat{p} \underset{\text{approx.}}{\sim} \text{Normal} \left(\mu_{\hat{p}} = p; \sigma_{\hat{p}}^2 = \frac{p(1-p)}{n} \right)$$

$$\text{So; } z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

* Likewise, if number of success is under consideration, we have

$$T \underset{\text{approx.}}{\sim} \text{Normal} (\mu = np; \sigma^2 = np(1-p))$$

and $P(T \geq b) = P\left(z \geq \frac{b - 0.5 - np}{\sqrt{np(1-p)}}\right)$ and

$$P(T \leq b) = P\left(z \leq \frac{b + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Since T is discrete and z is continuous, we have ± 0.5 .

6.27 The quality of computer disks is measured by the number of missing pulses. For a certain brand of disk, 80% are generally found to contain no missing pulses. If 100 such disks are inspected, find the approximate probability that 15 or fewer contain missing pulses.

6.27) T : # of disks that contain missing pulses.

$$T \sim \text{Binomial}(n=100; p=0,20) \\ = 1-0,80$$

$$P(T \leq 15) = P\left(\frac{T - np}{\sqrt{np(1-p)}} < \frac{15,5 - 100 \cdot 0,20}{\sqrt{100 \cdot 0,20 \cdot 0,80}}\right)$$

$$= P(Z < -1,13) = 0,5 - 0,3708 = 0,1292$$

6.28 The capacitances of a certain type of capacitor are normally distributed with a mean of 53 μf (microfarads) and a standard deviation of 2 μf . If 64 such capacitors are to be used in an electronic system, approximate the probability that at least 12 of them will have capacitances below 50 μf .

6.28) $X \sim \text{Normal}(\mu=53; \sigma^2=2^2)$

$$p = P(X < 50) = P\left(\frac{X - \mu}{\sigma} < \frac{50 - 53}{2}\right) = P(Z < -1,5)$$

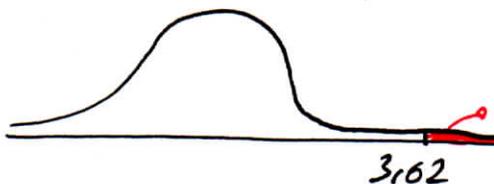
$$= 0,5 - 0,4332 = 0,0668$$

T : # of capacitors ^{below} ~~average~~ 50.

$$T \sim \text{Binomial}(n=64; p=0,0668)$$

$$P(T \geq 12) = P\left(\frac{T - np}{\sqrt{np(1-p)}} > \frac{11,5 - 64 \cdot 0,0668}{\sqrt{64 \cdot 0,0668 \cdot (1 - 0,0668)}}\right)$$

$$= P(Z > 3,62) \approx 0,0000$$



p can be ignored

$$(P(Z > 3) = 0,0001)$$

6.33 An auditor samples 100 of a firm's travel vouchers to check on how many of these vouchers are improperly documented. Find the approximate probability that more than 30% of the sampled vouchers will be found to be improperly documented if, in fact, only 20% of all the firm's vouchers are improperly documented.

6.33) $n = 100 ; p = 0,2$

$$\hat{p} \sim \text{Normal} \left(\mu_{\hat{p}} = 0,2 ; \sigma_{\hat{p}}^2 = \frac{0,2 \cdot 0,8}{100} = 0,0016 \right)$$

$$\sigma_{\hat{p}} = 0,04$$

$$P(\hat{p} > 0,3) = P\left(\frac{\hat{p} - p}{\sigma_{\hat{p}}} > \frac{0,3 - 0,2}{0,04}\right) = P(Z > 2,5) \\ = 0,5 - 0,4938 = 0,0062$$

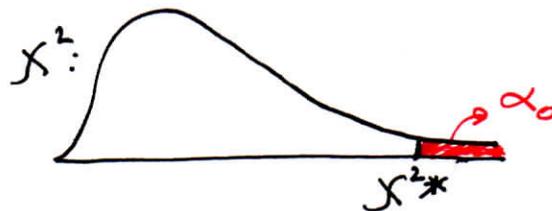
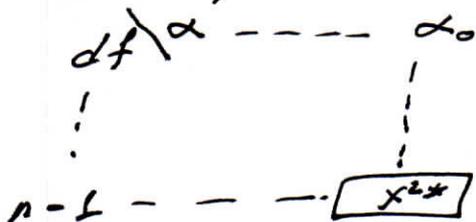
Sample Variance; S^2

* We used to find probabilities for \bar{X} and \hat{p} using $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ or $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$. For S^2 ,

we have a new distribution; $\chi^2 = \frac{(n-1) \cdot S^2}{\sigma^2}$.

* $n-1$ is degrees of freedom. χ^2 distribution does NOT have negative values and NOT symmetric. Moreover, the table gives χ^2 values, NOT probabilities.

we have;



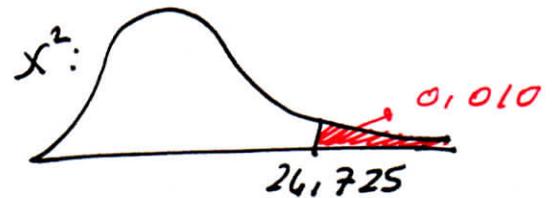
- 6.42 In constructing an aptitude test for a job, it is important to plan for a fairly large variance in test scores so the best applicants can be easily identified. For a certain test, scores are assumed to be normally distributed with a mean of 80 and a standard deviation of 10. A dozen applicants are to take the aptitude test. Find the approximate probability that the sample standard deviation of the scores for these applicants will exceed 15.
- 6.43 For an aptitude test for quality-control technicians in an electronics firm, history shows scores to be normally distributed with a variance of 225. If 20 applicants are to take the test, find an interval in which the sample variance of test scores should lie with probability 0.90.

$$6.42) X \sim \text{Normal} (\mu = 80; \sigma^2 = 10^2) \quad n = 12$$

$$P(s^2 > 15^2) = P\left(\frac{(n-1) \cdot s^2}{\sigma^2} > \frac{11 \cdot 15^2}{10^2}\right)$$

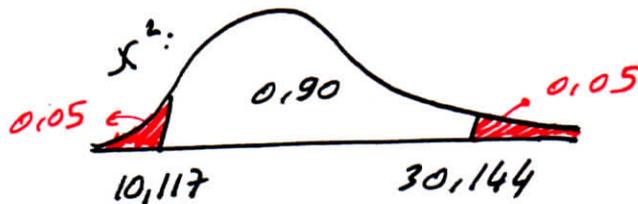
$$= P(\chi^2 > 24,75) \approx 0,010$$

(df = 11)



$$6.43) X \sim \text{Normal} (\mu; \sigma^2 = 225) \quad n = 20$$

df = n - 1 = 19



$$P(10,117 < \chi^2 < 30,144) = 0,90$$

$$10,117 < \frac{(n-1) \cdot s^2}{\sigma^2} < 30,144$$

$$10,117 < \frac{19 s^2}{225} < 30,144$$

$$\frac{225}{19} \cdot 10,117 < s^2 < 30,144 \cdot \frac{225}{19}$$

$$P(119,81 < s^2 < 356,97) = 0,90$$



* Let $U \sim \chi^2_{(n-1)}$ where $n-1$ stands for df.

We have, $E(U) = n-1$ and

$$\text{Var}(U) = 2(n-1)$$

Then, $E\left(\frac{(n-1)S^2}{\sigma^2}\right) = \frac{(n-1)}{\sigma^2} \cdot E(S^2) = n-1$

$$E(S^2) = \sigma^2 \text{ and}$$

$$\text{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) = \frac{(n-1)^2}{\sigma^4} \text{Var}(S^2) = 2 \cdot (n-1)$$

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

* We use these results in

(i) Finding a lower bound using ~~Tchebyscheff's~~ theorem $P(\mu - k\sigma \leq X \leq \mu + k \cdot \sigma) \geq 1 - \frac{1}{k^2}$

(ii) Unbiasedness property (Chapter 7) of S^2

Ex For Ex. 6.63, find ~~an~~ interval in which at least 75% of sample variances should lie.

Ans $\mu_{S^2} = \sigma^2$; $\sigma_{S^2}^2 = \frac{2\sigma^4}{n-1}$; $1 - \frac{1}{k^2} = 0,75$
 $k = 2$

The interval is $\mu_{S^2} \pm 2 \cdot \sigma_{S^2}$

Because S^2 cannot be negative

$$225 \pm 2 \cdot \sqrt{\frac{2 \cdot 225^2}{19}}$$

$$(0 ; 33074,84)$$