Q1) $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ by definition}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}X_i^2 - 2\bar{X}\underbrace{\sum_{i=1}^{n}X_i}_{n\bar{X}} + n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n}X_i^2 - n^2\bar{X}^2\right]$$

$$= \frac{1}{n(n-1)}\left[n\sum_{i=1}^{n}X_i^2 - \left(\sum_{i=1}^{n}X_i\right)^2\right] \text{ since } \sum_{i=1}^{n}X_i = n\bar{X}$$

(shown) #

Given: $\sum x_i^2 = 171$

$\sum x_i = 31$

$n = 6$

$$s^2 = \frac{1}{(6)(5)}\left[(6)(171) - (31)^2\right] = \frac{65}{30} = \frac{13}{6}$$

$\therefore$ variance of data is $\frac{13}{6}$ #

Q2 (i) $E[\bar{X}] = E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$

$$= \frac{1}{n}\left[\underbrace{E[X_1]}_{\mu} + \underbrace{E[X_2]}_{\mu} + \cdots + \underbrace{E[X_n]}_{\mu}\right]$$

$$= \frac{1}{\cancel{n}}\left[\cancel{n}\mu\right] = \mu \quad \text{(shown)}$$

(ii) $Var[\bar{X}] = Var\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]$

$$= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + \cdots + Var(X_n)\right]$$

Covariance between X_i and X_j $(i \neq j)$ all zero

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \cdots + \sigma^2\right]$$

$$= \frac{1}{n^{\cancel{2}}}\left[\cancel{n}\sigma^2\right]$$

$$= \frac{\sigma^2}{n}$$

<u>Remark</u>: This method is easier than using the definition $Var[\bar{X}] = E[(\bar{X} - \mu)^2]$ which we cover in class.

(iii) $$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} E\left[(X_i - \bar{X})^2\right] \quad \text{——— (1)}$$

We first consider $E\left[(X_i - \bar{X})^2\right]$; $i = 1, 2, \cdots, n$.

$$(X_i - \bar{X})^2 = X_i^2 - 2X_i\bar{X} + \bar{X}^2$$

$$= X_i^2 - 2X_i\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] + \left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right]^2$$

$$= X_i^2 - \frac{2}{n}X_i(X_1 + X_2 + \cdots + X_n) + \frac{1}{n^2}\left\{\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right\}$$

Taking expectation,

$E\left[(X_i - \bar{X})^2\right]$

$$= E\left[X_i^2\right] - \frac{2}{n}\left\{E\left[X_i^2\right] + \sum_{\substack{i \neq j \\ j=1}}^{n} E\left[X_i X_j\right]\right\} + \frac{1}{n^2}\left\{\sum_{i=1}^{n} E(X_i^2) + \sum_{\substack{i \neq j \\ i,j=1}}^{n} E\left[X_i X_j\right]\right\}$$

① Random variables are independent.
$E[X_i X_j] = E[X_i]\,E[X_j]$; $i \neq j$

② $Var(X_i) = E(X_i^2) - [E(X_i)]^2 \Rightarrow E(X_i^2) = \sigma^2 + \mu^2$

$$= (\sigma^2 + \mu^2) - \frac{2}{n}\left\{(\sigma^2 + \mu^2) + (n-1)\mu^2\right\} + \frac{1}{n^2}\left\{n(\sigma^2 + \mu^2) + (n^2 - n)\mu^2\right\}$$

$$= \left(\sigma^2 - \frac{2}{n}\sigma^2 + \frac{1}{n}\sigma^2\right) + \underbrace{\left(\mu^2 - 2\mu^2 + \mu^2\right)}_{0}$$

$$= \frac{n-1}{n}\sigma^2 \quad ; \quad i = 1, 2, \cdots, n$$

Substituting into (1),

$$E(S^2) = \frac{1}{\cancel{n-1}}(\cancel{n})\left[\frac{\cancel{n-1}}{\cancel{n}}\sigma^2\right) = \sigma^2 \quad \# \text{ (shown)}$$

<u>Alternative Solution</u>

We consider $Var(X_i - \bar{X})$ first.

$$Var(X_i - \bar{X}) = Var(X_i) + Var(\bar{X}) - 2\,Cov(X_i, \bar{X})$$

$$= \sigma^2 + \frac{\sigma^2}{n} - 2\,Cov\left(X_i, \frac{\sum_{i=1}^{n} X_i}{n}\right)$$

1) $Cov(X_i, X_i) = Var(X_i)$
2) $Cov(X_i, X_j) = 0$ by independence

$$= \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n} Var(X_i)$$

$$= \sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}\sigma^2 = \frac{n-1}{n}\sigma^2$$

$$Var(X_i - \bar{X}) = E\left[(X_i - \bar{X})^2\right] - \left[E(X_i - \bar{X})\right]^2$$

1) $Var(X_i - \bar{X}) = \frac{n-1}{n}\sigma^2$
2) $E(X_i - \bar{X}) = E(X_i) - E(\bar{X}) = \mu - \mu = 0$

$$\therefore\ E\left[(X_i-\bar{X})^2\right] = Var(X_i-\bar{X}) + \left[E(X_i-\bar{X})\right]^2$$

$$= \frac{n-1}{n}\sigma^2 + 0$$

$$= \frac{n-1}{n}\sigma^2 \qquad ; \ i = 1, 2, \cdots, n$$

$$E(S^2) = \frac{1}{n-1}\sum_{i=1}^{n} E\left[(X_i-\bar{X})^2\right]$$

$$= \frac{1}{\cancel{n-1}}\,(\cancel{n})\left[\frac{\cancel{n-1}}{\cancel{n}}\sigma^2\right] = \sigma^2 \quad \text{(shown)}$$

\#

(iv) $$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{S/\sigma}$$

$$= \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{S^2}{\sigma^2}}}$$

$$\text{let } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\text{let } V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Z and V are independent

$$= \frac{Z}{\sqrt{\dfrac{V}{n-1}}} \sim t_{n-1} \qquad \text{(shown)}$$

Q2(iii)

From Q1) $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$

Taking expectation,

$$E[S^2] = \frac{1}{n-1} E\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E[X_i^2] - nE[\bar{X}^2]\right]$$

$Var[X_i] = E[X_i^2] - (E[X_i])^2$

$\Rightarrow E[X_i^2] = \sigma^2 + \mu^2$

$Var[\bar{X}] = E[\bar{X}^2] - (E[\bar{X}])^2$

$\Rightarrow E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right]$$

$$= \frac{1}{n-1}\left[n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}\right]$$

$$= \frac{1}{\cancel{n-1}}\left[\cancel{(n-1)}\sigma^2\right]$$

$$= \sigma^2$$

#

Q3 (i) If X and Y are independent random variables such that

$$X \sim \chi^2_{n_1}$$
$$Y \sim \chi^2_{n_2}$$

then $$\frac{X/n_1}{Y/n_2} \sim F_{n_1, n_2}.$$

In our case, $X = \dfrac{V/n}{W/m}$ where $\left.\begin{array}{l} V \sim \chi^2_n \\ W \sim \chi^2_m \end{array}\right\}$ V, W independent

and $X \sim F_{n,m}$.

Now, $$X^{-1} = \frac{W/m}{V/n} \sim F_{m,n} \quad \# \quad \text{(shown)}$$

(ii) If Z and V are independent random variables such that

$$Z \sim N(0, 1^2)$$
$$V \sim \chi^2_n$$

then $$T = \frac{Z}{\sqrt{V/n}} \sim t_n$$

Now $T^2 = \dfrac{Z^2}{V/n}$ where $Z^2 \sim \chi_1^2$

$= \dfrac{Z^2/1}{V/n} \sim F_{1,n}$ # (shown).

Q4) Given: $n = 40$, $\mu = 28$, $\sigma = 5$

$X \sim$ unknown distribution

$\bar{X} \sim N(28, \frac{25}{40})$ by CLT

$P(\bar{X} > 30)$

$= P\left(\dfrac{\bar{X}-28}{\sqrt{25/40}} > \dfrac{30-28}{\sqrt{25/40}}\right)$

$= P(Z > 2.53)$

$= 1 - P(Z \leq 2.53)$

$= 1 - 0.9943$

$= 0.0057$

#

Q5) Given : $\sigma^2 = 1$

$X \sim$ unknown distribution

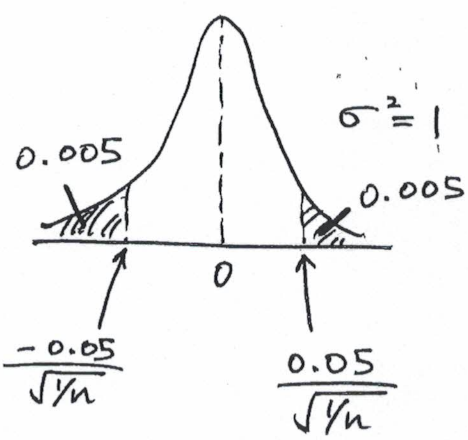
$\bar{X} \sim N(\mu, \frac{1}{n})$ by CLT

$$P(|\bar{X} - \mu| < 0.05) = 0.99$$

$$\Rightarrow P(-0.05 < \bar{X} - \mu < 0.05) = 0.99$$

$$\Rightarrow P\left(\frac{-0.05}{\sqrt{1/n}} < \frac{\bar{x} - \mu}{\sqrt{1/n}} < \frac{0.05}{\sqrt{1/n}}\right) = 0.99$$

$$\Rightarrow P\left(\frac{-0.05}{\sqrt{1/n}} < Z < \frac{0.05}{\sqrt{1/n}}\right) = 0.99$$ where $Z \sim N(0,1)$



From statistical table, $P(Z \leqslant 2.575) = 0.995$

Equating, $2.575 = \frac{0.05}{\sqrt{\frac{1}{n}}}$

$$\Rightarrow \frac{1}{n} = 3.770 \times 10^{-4}$$

$$\Rightarrow n = 2652.25 \approx 2653 \quad \#$$

I round up so that $\sqrt{\frac{1}{n}}$ decreases. As a result, the reqd probability is slightly bigger than 0.99

Q6) Given:

$\underline{A}$	$\underline{B}$
$\mu_A = 6.5$	$\mu_B = 6.0$
$\sigma_A = 0.9$	$\sigma_B = 0.8$
$n_A = 36$	$n_B = 49$

$X_A - X_B \sim$ unknown distribution.

$\overline{X}_A - \overline{X}_B \sim N\left(6.5 - 6.0, \frac{0.9^2}{36} + \frac{0.8^2}{49}\right)$ by CLT.

i.e. $N(0.5, 0.189^2)$

$P(\overline{X}_A - \overline{X}_B \geq 1.0)$

$= P\left(\frac{(\overline{X}_A - \overline{X}_B) - 0.5}{0.189} \geq \frac{1.0 - 0.5}{0.189}\right)$

$= P(Z \geq 2.65)$

$= 1 - P(Z < 2.65)$

$= 1 - 0.9960$

$= 0.0040$

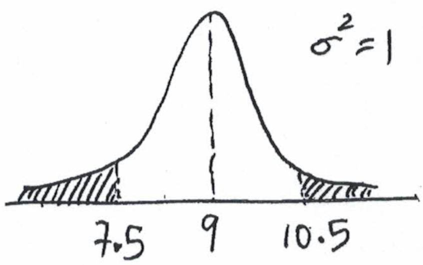
#

Q7) Given: $\mu = 9$, $\sigma = 1$, normality

Let X be weight of a carton.

$X \sim N(9, 1^2)$

(i)



Required region is shaded region.

$$\text{Proportion} = 2P(X > 10.5)$$
$$= 2(1 - P(X \leq 10.5))$$
$$= 2\left(1 - P\left(\frac{X-9}{1} \leq \frac{10.5-9}{1}\right)\right)$$
$$= 2(1 - P(Z \leq 1.5))$$
$$= 2(1 - 0.9332)$$
$$= 0.1336$$

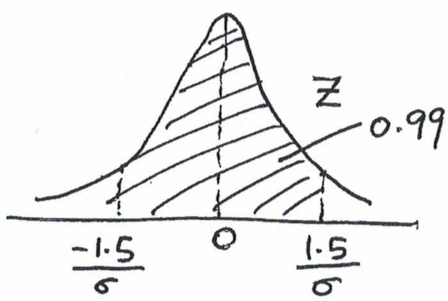
#

(ii) Now $X \sim N(9, \sigma^2)$. We want

$$P(7.5 \leq X \leq 10.5) = 0.99$$

$$\Rightarrow P\left(\frac{7.5-9}{\sigma} \leq \frac{X-9}{\sigma} \leq \frac{10.5-9}{\sigma}\right) = 0.99$$

$$\Rightarrow P\left(\frac{-1.5}{\sigma} \leq Z \leq \frac{1.5}{\sigma}\right) = 0.99$$



$\therefore$ By symmetry, $P\left(Z \leq \frac{1.5}{\sigma}\right) = 0.995$

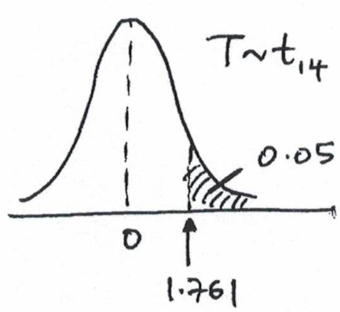
From statistical table, $\frac{1.5}{\sigma} = 2.575$

$$\Rightarrow \sigma = \frac{1.5}{2.575} \approx 0.5825$$

#

Q8) Given: $n = 15$, $T \sim t_{14}$

From statistical table, $P(T \geq 1.761) = 0.05$



By symmetry,

$$P(T < -1.761) = 0.05 \quad \text{—(1)}$$

Also, $P(k < T < -1.761)$

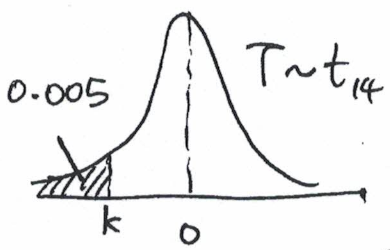
$= P(T < -1.761) - P(T \leq k)$

So we have

$$P(T < -1.761) - P(T \leq k) = 0.045 \quad \text{—(2)}$$

Substituting (1) into (2):

$$P(T \leq k) = 0.05 - 0.045 = 0.005$$



By symmetry,

$$P(T \geqslant -k) = 0.005$$

From statistical table, $-k = 2.977$

Hence $k = -2.977$

#

Q9) Given: $n_1 = 25$, $\sigma_1^2 = 10$, normality | $n_2 = 31$, $\sigma_2^2 = 15$, normality

$$P\left(\frac{S_1^2}{S_2^2} > 1.26\right) = P\left(\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} > \left(\frac{\sigma_2^2}{\sigma_1^2}\right) 1.26\right)$$

$$= P(W > 1.89)$$ where $W \sim F_{24,30}$

$$= 0.05$$ from statistical table

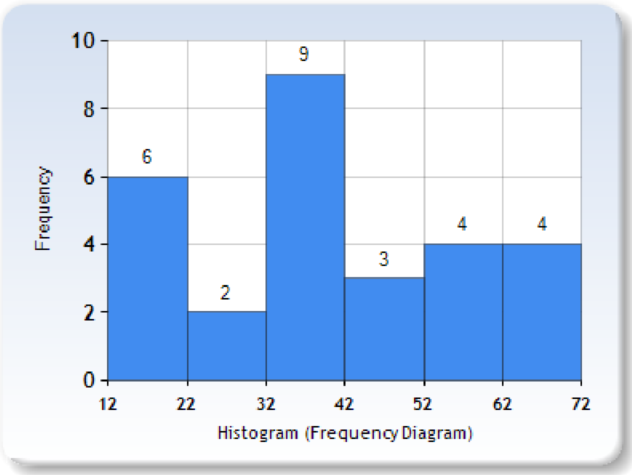
#

10)

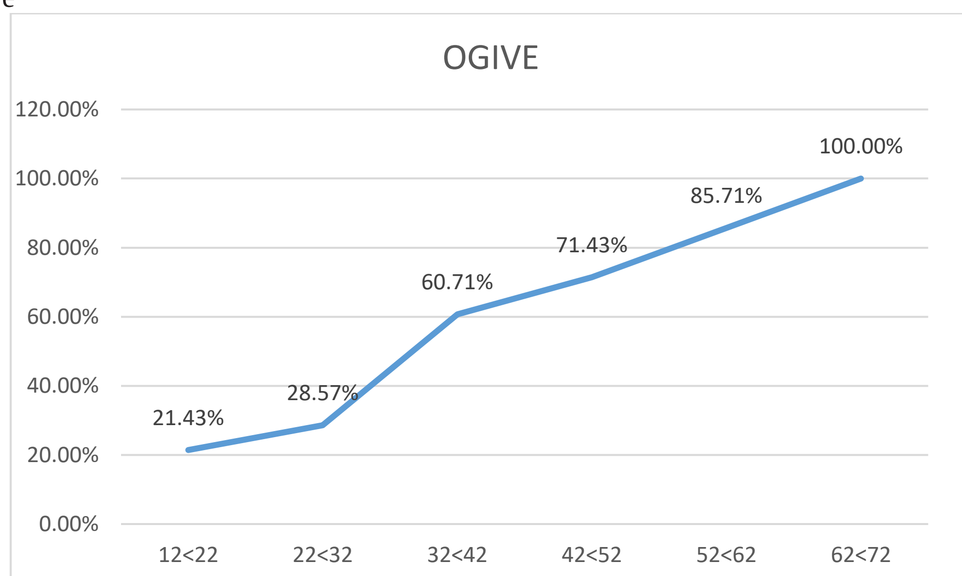
a. frequency distribution

Class	Frequency
12<22	6
22<32	2
32<42	9
42<52	3
52<62	4
62<72	4

b. histogram



c. ogive



d. stem-and-leaf display

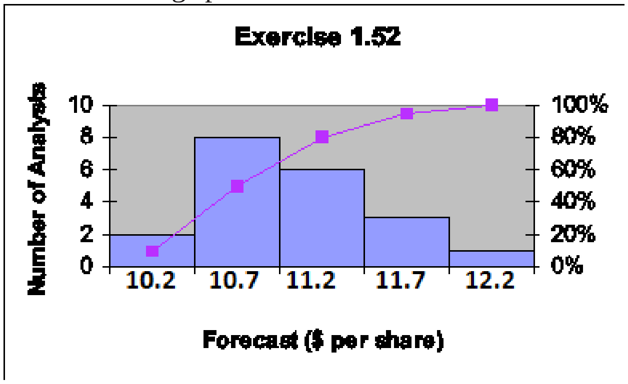
S	L
1	23557
2	148
3	2567799
4	0144
5	14699
6	2455

11)

Classes	Frequency	a. Relative Frequency	b. Cumulative Frequency	c. Cumulative Relative Frequency
0<10	8	16.33%	8	16.33%
10<20	10	20.41%	18	36.74%
20<30	13	26.53%	31	63.27%
30<40	12	24.49%	43	87.76%
40<50	6	12.24%	49	100.00%
Total	49	100.00%		

12)

a. Draw a histogram of 20 forecasted earnings per share.



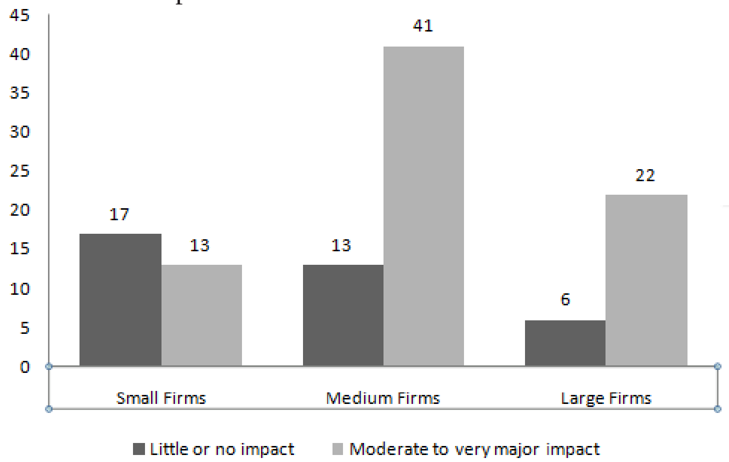
Answer to b., c. and d. are:

	(b)	(c)	(d)
Frequency	Relative Freq.	Cumulative Freq.	Cumulative %
2	0.1	2	10.00%
8	0.4	10	50.00%
6	0.3	16	80.00%
3	0.15	19	95.00%
1	0.05	20	100.00%

d. Cumulative relative frequencies are in the last column of the table above. These numbers indicate the percent of analysts who forecast that level of earnings per share and all previous classes, up to and including the current class. The third bin of 80% indicates that 80% of the analysts have forecasted up to and including that level of earnings per share.

13)

Cluster bar chart for impact of SOX



14)

a. Cross table

Type of Account	Male	Female	Subtotal
Easy Checking	80	100	180
Intelligent Checking	12	24	36
Super Checking	27	27	54
Ultimate Checking	24	6	30
Subtotal	143	157	300

b. Stacked bar chart