1.

(a) $E(X)=(10000)(3.848\times10^{-7})+(5000)(1.501\times10^{-5})+\cdots+(10)(0.4226)=7.0988$

$E(X^2)=(10000)^2(3.848\times10^{-7})+(5000)^2(1.501\times10^{-5})+\cdots+(10)^2(0.4226)=994.1455$

$Var(X)=E(X^2)-E(X)^2=943.7519$

(b) $Y=X-10$

$E(Y)=E(X)-10=-2.9012$

$Var(Y)=Var(X)=943.7519$

(c) Since expected gain is negative, this game is not a fair game.

(d) $p=1-3.848\times10^{-7}-1.501\times10^{-5}-\cdots-0.4226=0.5011$

(e) $W\sim b(10,0.5011)$

(f) Using normal approximation with continuity correction,

$$\Pr(W>50)=\Pr\left(\frac{W-(100)(0.5011)}{\sqrt{(100)(0.5011)(0.4989)}}\geq\frac{50.5-(100)(0.5011)}{\sqrt{(100)(0.5011)(0.4989)}}\right)$$
$$\approx1-\Phi(0.08)=1-0.5319=0.4681$$

(g) Let Y_i be the money you win in ith game, then the overall gain after the 100 games is $\sum_{i=1}^{100}Y_i$. Hence

$$\Pr\left(\sum_{i=1}^{100}Y_i<0\right)=\Pr(\bar{Y}<0)=\Pr\left(\frac{\bar{Y}-E(Y)}{\sqrt{Var(Y)/100}}<\frac{0-(-2.9012)}{\sqrt{943.7519/100}}\right)$$
$$\approx\Phi(0.94)=0.8264 \qquad \text{(normal approximation)}$$

2.

(a) $$E(X)=\sum_{i=1}^{K}i\Pr(X=i)=\frac{1}{K}\sum_{i=1}^{K}i=\frac{1}{K}\frac{K(K+1)}{2}=\frac{K+1}{2}$$

$$E(X^2)=\sum_{i=1}^{K}i^2\Pr(X=i)=\frac{1}{K}\sum_{i=1}^{K}i^2=\frac{1}{K}\frac{K(K+1)(2K+1)}{6}=\frac{(K+1)(2K+1)}{6}$$

$$Var(X)=E(X^2)-E(X)^2=\frac{K(K+1)(2K+1)}{6}-\frac{(K+1)^2}{4}=\frac{K^2-1}{12}$$

(b) Method of moment estimator of K is given by :

$$\bar{X}=\frac{\hat{K}+1}{2}\Leftrightarrow\hat{K}=2\bar{X}-1$$

(c) From the data, $\bar{X}=24.88$, hence $\hat{K}=2(24.88)-1=48.76\approx49$

(d) Stem-and-leaf plot for this sample :

```
0*|24788
1 |03589            n = 25
2 |11678            leaf unit = 1
3 |24479
4 |11377
```

(e) Lower quartile is the $0.25\times(25+1)$ th number in ordered sample, i.e.

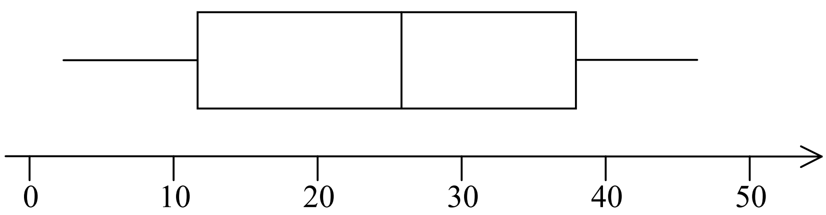
$$Q_1 = X_{(6.5)} = \frac{1}{2}\left(X_{(6)} + X_{(7)}\right) = \frac{1}{2}(10+13) = 11.5$$

Similarly,

$$Median = X_{(13)} = 26$$

$$Q_2 = X_{(19.5)} = \frac{1}{2}\left(X_{(19)} + X_{(20)}\right) = \frac{1}{2}(37+39) = 38$$

Box-plot :



3.

(a) Let X be the number of cars passing the point within 5 minutes, then $X \sim \wp(5)$.

$$\Pr(X=6) = \frac{e^{-5}5^6}{6!} = 0.1462$$

(b) Let Y be the number of cars passing the point within one hour, then $Y \sim \wp(60)$.
Use normal approximation with continuity correction,

$$\Pr(Y \ge 70) = \Pr\left(\frac{Y-60}{\sqrt{60}} \ge \frac{69.5-60}{\sqrt{60}}\right) \approx 1-\Phi(1.23) = 1-0.8907 = 0.1093$$

(c) Let T be the waiting time for 70 cars passing the point, then

$$T \sim \Gamma(70,1) \Leftrightarrow 2T \sim \Gamma\left(70,\frac{1}{2}\right) \equiv \chi^2_{140}$$

Use normal approximation,

$$\begin{aligned}\Pr(T>60) &= \Pr(2T>120)\\ &= \Pr\left(\frac{2T-140}{\sqrt{280}} > \frac{120-140}{\sqrt{280}}\right) \approx 1-\Phi(-1.20) = 0.8849\end{aligned}$$

4.

(a) Let μ_x, σ_x^2 be the population mean and variance of the computing time by design 1; μ_y, σ_y^2 be the population mean and variance of the computing time by design 2. Then the hypotheses are :

$$H_0 : \mu_x = \mu_y \qquad \text{vs} \qquad H_1 : \mu_x \neq \mu_y$$

(b) Test $H_0 : \mu_x = \mu_y$ vs $H_1 : \mu_x \neq \mu_y$ at $\alpha = 0.05$.

Test statistics : $T = \dfrac{\overline{X} - \overline{Y}}{S_{pool}\sqrt{\dfrac{1}{16} + \dfrac{1}{13}}}$

Reject H_0 at $\alpha = 0.05$ if $|T| > t_{27,0.025} = 2.052$.

From the data, $\overline{X} = 2.2$, $S_x^2 = 0.1427$, $\overline{Y} = 2.5154$, $S_y^2 = 0.1231$.

$$S_{pool}^2 = \frac{(16-1)(0.1427) + (13-1)(0.1231)}{16+13-2} = 0.1340$$

$$T_{obs} = \frac{2.2 - 2.5154}{\sqrt{(0.1340)\left(\dfrac{1}{16} + \dfrac{1}{13}\right)}} = -2.3075 \Rightarrow |T_{obs}| = 2.3075 > 2.052$$

Hence reject H_0 at $\alpha = 0.05$. The data had provided strong evidence that the two design produces different average computing time.

(c) Test $H_0 : \sigma_x^2 = \sigma_y^2$ vs $H_1 : \sigma_x^2 \neq \sigma_y^2$ at $\alpha = 0.1$.

Test statistic : $F = \dfrac{S_x^2}{S_y^2}$

Reject H_0 at $\alpha = 0.1$ if $F > F(15,12,0.05) = 2.62$ or $F < F(15,12,0.95) = 0.4032$.

From the data, $F_{obs} = \dfrac{0.1427}{0.1231} = 1.1592 \Rightarrow 0.0432 < F_{obs} < 2.62$.

Hence do not reject H_0 at $\alpha = 0.1$. The data didn't show strong evidence that the equal variance assumption is violated.

(d) A 95% confidence interval for μ_x is given by

$$\overline{X} \pm t_{15,0.025}\frac{S_x}{\sqrt{16}} = 2.2 \pm (2.131)\sqrt{\frac{0.1427}{16}} = 2.2 \pm 0.2012 = [1.9988\,, 2.4012]$$

(OR $\overline{X} \pm t_{27,0.025}\dfrac{S_{pool}}{\sqrt{16}} = 2.2 \pm (2.052)\sqrt{\dfrac{0.1340}{16}} = 2.2 \pm 0.1878 = [2.0122\,, 2.3878]$)

A 95% confidence interval for μ_y is given by

$$\overline{Y} \pm t_{12,0.025}\frac{S_y}{\sqrt{13}} = 2.5154 \pm (2.179)\sqrt{\frac{0.1231}{13}} = 2.5154 \pm 0.2120 = [2.3034\,, 2.7274]$$

(OR $\overline{Y} \pm t_{27,0.025}\dfrac{S_{pool}}{\sqrt{13}} = 2.5154 \pm (2.052)\sqrt{\dfrac{0.1340}{13}} = 2.5154 \pm 0.2083 = [2.3071\,, 2.2.7237]$)

(e) A 90% confidence interval for $\mu_x - \mu_y$ is given by

$$\left(\overline{X}-\overline{Y}\right)\pm t_{27,0.025}S_{pool}\sqrt{\frac{1}{16}+\frac{1}{13}} = (2.2-2.5154)\pm(2.052)\sqrt{(0.1340)\left(\frac{1}{16}+\frac{1}{13}\right)}$$
$$= -0.3154 \pm 0.2805 = [-0.5959\,, -0.0349]$$

5.

(a) Test $H_0 : p_M = p_F$ vs $H_1 : p_M \neq p_F$ at $\alpha = 0.05$.

Test statistic : $Z = \dfrac{\widehat{p}_M - \widehat{p}_F}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{300}+\frac{1}{300}\right)}}$

Reject H_0 at $\alpha = 0.05$ if $|Z| > Z_{0.025} = 1.96$.

From the data, $\widehat{p}_M = \dfrac{180}{300} = 0.6$, $\widehat{p}_F = \dfrac{125}{300} = 0.4167$, $\widehat{p} = \dfrac{180+125}{300+300} = 0.5083$.

$$Z_{obs} = \frac{0.6-0.4167}{\sqrt{(0.5083)(0.4917)\left(\frac{1}{300}+\frac{1}{300}\right)}} = 4.4905 \Rightarrow |Z_{obs}| > 1.96$$

Hence reject H_0 at $\alpha = 0.05$.

(b) The p-value of the test in (a) is given by

$$p-value = 2\Pr(Z > 4.4905 \mid H_0) \approx 0$$

(c) A 95% confidence interval for $p_M - p_F$ is given by

$$(\widehat{p}_M - \widehat{p}_F) \pm Z_{0.025}\sqrt{\frac{\widehat{p}_M(1-\widehat{p}_M)}{300}+\frac{\widehat{p}_F(1-\widehat{p}_F)}{300}}$$
$$= (0.6-0.4167)\pm(1.96)\sqrt{\frac{(0.6)(0.4)}{300}+\frac{(0.4167)(0.5833)}{300}}$$
$$= 0.1833 \pm 0.07865 = [0.10465\,, 0.26195]$$

Since the whole interval is on the right hand side of zero, we have high confidence that p_M is greater than p_F, i.e. the recruitment is biased towards male applicants.

(d) Using the estimates from last year, the sample sizes needed are :

$$n_M = \frac{Z_{0.025}^2\,\widehat{p}_M(1-\widehat{p}_M)}{D^2} = \frac{(1.96)^2(0.6)(0.4)}{(0.05)^2} = 368.7936 \approx 369$$
$$n_F = \frac{Z_{0.025}^2\,\widehat{p}_F(1-\widehat{p}_F)}{D^2} = \frac{(1.96)^2(0.4167)(0.5833)}{(0.05)^2} = 373.4974 \approx 374$$

OR, if you use the conservative bound for $p(1-p)$, then

$$n_M = n_F = \frac{Z_{0.025}^2}{4D^2} = \frac{(1.96)^2}{4(0.05)^2} = 384.16 \approx 385$$