1.(a) Denote the sample proportion as $\hat{p} = X/n = X/150$, the estimator of p. Then the test statistic for H_0: $p = 0.2$ is $Z = (\hat{p} - 0.2)/\sqrt{.2(.8)/150} = (\hat{p} - 0.2)/0.03266$, and we reject H_0 in favor of H_1 at level $\alpha = 0.05$ when $Z > z_{.05} = 1.645$; equivalently, reject when $\hat{p} > 0.2 + (1.645)(0.03266) = 0.253725$.

(b) When $p = 0.25$ is the true value, $\hat{p}$ has approximate normal distribution with mean $E(\hat{p}) = 0.25$ and St.Dev$(\hat{p}) = \sqrt{.25(.75)/150} = 0.035355$. Thus, since from (a) we do not reject when $\hat{p} < 0.253725$, we have

$$\beta = P(\hat{p} < 0.253725 \mid p = 0.25) = P\left(\frac{\hat{p} - 0.25}{0.035355} < \frac{0.253725 - 0.25}{0.035355}\right)$$

$$= P\big(N(0,1) < 0.105359\big) = \Phi(0.105) = 0.5428$$

(c) Observed $X = 39$ gives $\hat{p} = 39/150 = 0.26$, so the test statistic value is $Z_{obs} = (0.26 - 0.2)/0.03266 = 1.837 > 1.645$, so we reject H_0 in favor of H_1 at $\alpha = 0.05$, and conclude that $p > 0.2$. For the observed test statistic value $Z_{obs} = 1.837$, we have $P-\text{value} = P(Z > 1.837) = 1 - P(Z < 1.837) = 1 - .9668 = 0.0332 < .05$, consistent with rejecting H_0 at $\alpha = .05$ level.

(d) A 95% CI for p is

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} = 0.26 \pm 1.96\sqrt{.26(.74)/150} \equiv 0.26 \pm 1.96\,(0.0358) = 0.26 \pm 0.070,$$

so 0.19 to 0.33 is a 95% CI for p.

2.(a) We have $\hat{p}_1 = 39/150 = 0.26$ and $\hat{p}_2 = 58/200 = 0.29$, and a 95% CI for $p_1 - p_2$ is $(\hat{p}_1 - \hat{p}_2) \pm 1.96\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}$. Thus, the CI is

$$(0.26 - 0.29) \pm 1.96\sqrt{.26(.74)/150 + .29(.71)/200} = -0.03 \pm 1.96(0.048) = -0.03 \pm 0.094$$

so -0.124 to 0.064 is 95% CI for $p_1 - p_2$. Since this CI clearly includes 0, $p_1 - p_2 = 0$ is a very plausible value, so no evidence of any difference. (Additionally, can calculate test statistic for H_0: $p_1 - p_2 = 0$, find $Z = -0.62$, and $|Z| < 1.96$, so no evidence to reject H_0.)

(b) By the CLT, $\bar{X}_1$ and $\bar{X}_2$ have approximate normal distributions, so also does $\bar{X}_1 + \bar{X}_2$. The mean and variance of $\bar{X}_1 + \bar{X}_2$ are $E[\bar{X}_1 + \bar{X}_2] = E[\bar{X}_1] + E[\bar{X}_2] = \lambda_1 + \lambda_2 = 3.5 + 5.0 = 8.5$, and by independence, $\text{Var}[\bar{X}_1 + \bar{X}_2] = \text{Var}[\bar{X}_1] + \text{Var}[\bar{X}_2] = \lambda_1/n_1 + \lambda_2/n_2 = 3.5/40 + 5.0/50 = 0.1875$. Therefore, the standardized rv $Z = (\bar{X}_1 + \bar{X}_2 - 8.5)/\sqrt{0.1875} = (\bar{X}_1 + \bar{X}_2 - 8.5)/0.433$ has an approximate $N(0,1)$ distribution. Therefore,

$$P(\bar{X}_1 + \bar{X}_2 < 9.0) = P\left(\frac{\bar{X}_1 + \bar{X}_2 - 8.5}{0.433} < \frac{9.0 - 8.5}{0.433}\right) = P(Z < 1.155) = \Phi(1.155) = 0.876$$

(c) The pdf of a single rv X_i is $f(x_i;\theta) = \theta\, x_i^{\theta-1}$, so the likelihood function (joint pdf of $X_1, \ldots, X_n$) is

$$L(\theta) = f(x_1;\theta) \times f(x_2;\theta) \times \cdots \times f(x_n;\theta) = \theta x_1^{\theta-1}\, \theta x_2^{\theta-1} \cdots \theta x_n^{\theta-1} = \theta^n\, (x_1\, x_2 \cdots x_n)^{\theta-1}$$

for $\theta > 0$. The log of the likelihood function is then $l(\theta) = \ln[L(\theta)] = n\ln(\theta) + (\theta - 1)\ln(x_1\, x_2 \cdots x_n) \equiv n\ln(\theta) + (\theta-1)\sum_{i=1}^{n}\ln(x_i)$. We then obtain the likelihood equation,

$$\frac{\partial l}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln(x_i) = 0$$

with solution $\hat{\theta} = -n / \sum_{i=1}^{n} \ln(X_i)$, which is the MLE of θ.

(d) The cdf of X is $F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(u)\,du = \int_0^x \theta u^{\theta-1} du = u^\theta \Big|_0^x = x^\theta$, for $0 \le x \le 1$.
Cdf of Y is $F_Y(y) = P(Y \le y) = P(-\ln(X) \le y) = P(X \ge e^{-y}) = 1 - F_X(e^{-y}) = 1 - e^{-\theta y}$, $y > 0$.

3.(a) Point estimates of $\mu_1 - \mu_2$, σ^2, and St.Dev.$(\bar{X}_1 - \bar{X}_2) = \sigma\sqrt{1/11 + 1/16}$, respectively, are $\bar{X}_1 - \bar{X}_2 = 14.0 - 14.8 = -0.8$, $S_p^2 = [(11-1)\,10.5 + (16-1)\,8.0]/(11+16-2) = 9.0$, and $S_p\sqrt{1/11 + 1/16} = 1.175$.

(b) Since $t_{.025,25} = 2.060$, a 95% CI for $\mu_1 - \mu_2$ is

$$(\bar{X}_1 - \bar{X}_2) \pm 2.06\, S_p \sqrt{1/11 + 1/16} = -0.8 \pm 2.06\,(1.175) = -0.8 \pm 2.42$$

so -3.22 to 1.62 is a 95% CI for $\mu_1 - \mu_2$. A 95% CI for μ_2 is

$$\bar{X}_2 \pm t_{.025,15}\, S_2/\sqrt{n_2} = 14.8 \pm 2.131\sqrt{8.0/16} \equiv 14.8 \pm 1.507$$

so 13.293 to 16.307 is a 95% CI for μ_2.

(c) Since the 95% CI for $\mu_1 - \mu_2$ has the value 0 included well within the interval, we have no indication or evidence of any real difference in means under the two processes.
Formally, we can set up the test of $H_0\!: \mu_1 - \mu_2 = 0$ versus $H_1\!: \mu_1 - \mu_2 \ne 0$ at $\alpha = 0.05$ level; it uses the "pooled" t-statistic,

$$t = (\bar{X}_1 - \bar{X}_2)/S_p\sqrt{1/n_1 + 1/n_2} = -0.8/1.175 = -0.68\,,$$

which is well within the 'non-rejection' region ($-2.06 < t < 2.06$); so we would not reject H_0. No evidence to conclude there is any difference in means between the two processes.

(d) A large-sample 95% CI for $\mu_1 - \mu_2$ will be $(\bar{X}_1 - \bar{X}_2) \pm 1.96\, S_p\sqrt{1/n + 1/n}$, and with $S_p \approx \sigma = 3.0$, so we want n such that

$$e = 1.0 = 1.96\,(3.0)\sqrt{2/n}$$

Solving for n we get $n = (1.96)^2\,(3.0)^2\,2 \approx 69$.

4.(a) The data arise from a paired sampling design experiment, so the appropriate analysis to examine for mean difference in treatments (brands) is based on analysis of the sample of $n = 6$ differences in the data pairs, $D_i = X_{Ai} - X_{Bi}$. These differences D_i have values $2, 0, 3, 3, -1, 5$, with sample mean $\bar{D} = 12/6 = 2.0$ and sample variance $S_D^2 = 24/(6-1) = 4.8$, so $S_D = \sqrt{4.8} \approx 2.19$. Now since $t_{.05,5} = 2.015$, to test $H_0\!: \mu_D = 0$ versus $H_1\!: \mu_D \ne 0$ at level $\alpha = .10$, we use the test statistic (one-sample t-statistic) $t = \bar{D}/(S_D/\sqrt{n})$ and reject H_0 when $|t| > 2.015$. We have the test statistic value $t = 2.0/(2.19/\sqrt{6}) = 2.236 > 2.015$, so we reject H_0 at level $.10$, and conclude that the two brands do differ in mean wear (brand B has lower mean wear than A).

(b) A 90% CI for the mean difference $\mu_D = E(D_i)$ is $\bar{D} \pm t_{.05,5}\, S_D/\sqrt{n}$, so $2.0 \pm 2.015\,(2.19/\sqrt{6}) \equiv 2.0 \pm 1.802$, hence 0.198 to 3.802 is a 90% CI for μ_D. (Note this CI does not include the value $\mu_D = 0$, consistent with rejecting H_0 at level $\alpha = .10$.)

(c) (i) Should *randomize* assignment of shoe brand between left and right foot for runners, to prevent bias due to any possible systematic difference between wear of left and right foot.
(ii) Main reason is to reduce variance of estimate $\bar{D}$ of mean difference, by 'eliminating' source of variability due to differences in wear across runners.