

1. Suppose that the machine filling mini-boxes of Fruitland Raisins fills boxes so that the weight of the boxes has a population mean $\mu_x = 14.1$ grams and a standard deviation $\sigma_x = 1.4$ grams. Suppose that a random sample of 49 boxes is selected. What is the probability that the sample mean will be above 14.0 grams? There is a 99% chance that the sample mean will be above what value? What approximation are you making to answer these questions and why is it valid?

Since $n = 49 > 30$, by the rule of thumb, we may assume that the sample mean is normally distributed $\bar{X} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$. Thus

$$\begin{aligned} P(\bar{X} > 14.0) &= 1 - P(\bar{X} \leq 14.0) \\ &= 1 - P\left(Z = \frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}} \leq \frac{14.0 - 14.1}{1.4/\sqrt{49}} = -.5\right) \\ &= 1 - \Phi(-.5) = 1 - .3085 = \boxed{.6915}. \end{aligned}$$

To find x_{cr} so that with 99% probability the sample mean is above this number, solve

$$\begin{aligned} .99 &= P(\bar{X} > x_{\text{cr}}) = 1 - P(\bar{X} \leq x_{\text{cr}}) \\ &= 1 - P\left(Z = \frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}} \leq \frac{x_{\text{cr}} - \mu_x}{\sigma_x/\sqrt{n}}\right) \\ &= 1 - \Phi\left(\frac{x_{\text{cr}} - \mu_x}{\sigma_x/\sqrt{n}}\right) \end{aligned}$$

Remember that $.01 = P(Z \leq -z_{.01})$ so that

$$.01 = \Phi\left(\frac{x_{\text{cr}} - \mu_x}{\sigma_x/\sqrt{n}}\right) = \Phi(-z_{.01}).$$

Hence,

$$\frac{x_{\text{cr}} - \mu_x}{\sigma_x/\sqrt{n}} = -z_{.01}$$

so

$$x_{\text{cr}} = \mu_x - z_{.01} \frac{\sigma_x}{\sqrt{n}} = 14.1 - (2.326) \frac{1.4}{\sqrt{49}} = \boxed{-13.6}.$$

The approximation used is that $\bar{X} \sim N(\mu_x, \frac{\sigma_x}{\sqrt{n}})$ because of the Central Limit Theorem, which according to the rule of thumb is acceptable for $n > 30$.

2. Let $X \in \{0, 1, 2, 3\}$ be the number of cars and $Y \in \{0, 1, 2\}$ be the number of pickups in the Big Ed's parking lot at noon. The joint pmf is given by $p(x, y)$. What is the probability that $X \leq Y$? What is the expected number of pickups $E(Y)$? Given that $X = 1$, determine the conditional pmf of Y — i.e., find $p_{Y|X}(0|1)$, $p_{Y|X}(1|1)$, $p_{Y|X}(2|1)$. Given that one car is in the lot, what is the probability that there is at least one pickup? Are the number of cars X and the number of pickups Y independent r.v.'s?

$p(x, y)$		x			
		0	1	2	3
y	0	0	.1	.1	.2
	1	.1	.3	0	0
	2	.2	0	0	0

The event that $X \leq Y$ consists of the pairs $\{(0, 0), (0, 1), (0, 2), (1, 1), (1, 2), (2, 2)\}$ whose probability is

$$P = p(0, 0) + p(0, 1) + p(0, 2) + p(1, 1) + p(1, 2) + p(2, 2) = 0 + .1 + .2 + .3 + 0 + 0 = \boxed{0.6}.$$

The marginal probability for Y is gotten by summing rows, so $p_Y(0) = p_Y(1) = .4$ and $p_Y(2) = .2$. Thus the expected number of pickups is

$$E(Y) = \sum_{y=0}^2 y p_Y(y) = (0)(.4) + (1)(.4) + (2)(.4) = \boxed{0.8}.$$

The marginal probability for X is gotten by summing columns, so $p_X(0) = .3$, $p_X(1) = .4$, $p_X(2) = .1$ and $p_X(3) = .2$. The conditional probability is given by the formula

$$p_{Y|X}(y|x) = \frac{p(x, y)}{p_X(x)} \quad \text{so} \quad p_{Y|X}(0|1) = \frac{p(1, 0)}{p_X(1)} = \frac{.1}{.4} = \boxed{.25},$$

$$p_{Y|X}(1|1) = \frac{p(1, 1)}{p_X(1)} = \frac{.3}{.4} = \boxed{.75}, \quad p_{Y|X}(2|1) = \frac{p(1, 2)}{p_X(1)} = \frac{0}{.4} = \boxed{0}.$$

Thus,

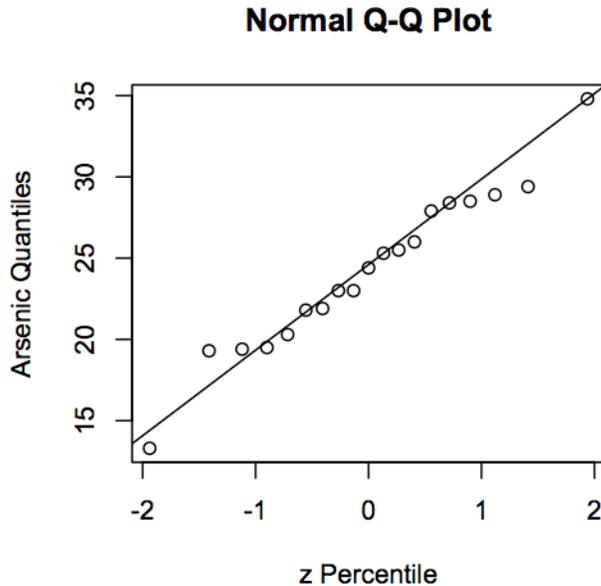
$$P(Y \geq 1|X = 1) = p_{Y|X}(1|1) + p_{Y|X}(2|1) = .75 + 0 = \boxed{0.75}.$$

Finally, X and Y are not independent, because, e.g.,

$$p_X(0)p_Y(0) = (.3)(.4) = .12 \neq 0 = p(0, 0).$$

3. In a study of poisons in the environment, the following measurements were made of ground-water arsenic in $\mu\text{g/L}$ at a certain landfill. The sample mean and standard deviation as found by **R** are $\bar{X} = 24.24$ and $s = 4.91$. Find a two sided 90% confidence interval for μ , the population mean arsenic level. What did you assume about the data in answering (a.)? Using the probability plot generated by **R**, comment on the validity of your assumptions.

25.5 24.4 19.5 20.3 28.4 23.0 34.8 29.4 25.3 19.4
 26.0 21.9 28.9 27.9 21.8 19.3 23.0 13.3 28.5



In this case, there are $n = 19$ observations, which is a small sample. Hence we use a CI based on the T -distribution. There are $\nu = n - 1 = 18$ degrees of freedom. Thus for the 90% confidence level, $\alpha = .10$ and the two sided critical value is $t_{\alpha/2, \nu} = t_{.05, 18} = 1.734$ from Table A.5. The two-sided CI is thus

$$\bar{X} \pm t_{\alpha/2, \nu} \frac{s}{\sqrt{n}} = 24.24 \pm (1.734) \frac{4.91}{\sqrt{19}} = \boxed{(22.29, 26.19)}.$$

The data is assumed to be approximately normally distributed. The points line up nicely in the probability plot, so the assumption that the data is approximately normal is reasonable. (In fact, a normal random number generator was used to generate this data. For small n , this probability plot is as straight as it gets for normal data.)

4. Suppose that reaction times are uniformly distributed in the interval $[0, c]$. Let X, Y be a random sample of reaction times. Consider two statistics $\hat{\theta}_1$ and $\hat{\theta}_2$. Show that $\hat{\theta}_1$ is an unbiased estimator for c . Show that $\hat{\theta}_2$ is an unbiased estimator for c . Find the standard error of $\hat{\theta}_1$. [The standard error for $\hat{\theta}_2$ is $\sigma_{\hat{\theta}_2} = \frac{c}{3\sqrt{2}}$.] Which is the better estimator for c ? Why?

$$\begin{aligned}\hat{\theta}_1 &= X + Y, \\ \hat{\theta}_2 &= \frac{3}{2} \max\{X, Y\}.\end{aligned}$$

For the uniform distribution on the interval $[0, c]$, the pdf $f(x) = 1/c$ if $0 \leq x \leq c$ and $f(x) = 0$ otherwise. The mean is

$$\mu = \int_{\mathbb{R}} x f(x) dx = \int_0^c \frac{x dx}{c} = \frac{c}{2},$$

and the expected square and variance is

$$E(X^2) = \int_0^c \frac{x^2 dx}{c} = \frac{c^2}{3}, \quad \sigma^2 = E(X^2) - E^2(X) = \frac{c^2}{3} - \frac{c^2}{4} = \frac{c^2}{12}.$$

X and Y are independent variables taken from this distribution. $\hat{\theta}_1$ is unbiased since

$$E(\hat{\theta}_1) = E(X + Y) = E(X) + E(Y) = \frac{c}{2} + \frac{c}{2} = c.$$

We note that $\hat{\theta}_2 = X$ if $X \geq Y$ and $\hat{\theta}_2 = Y$ if $X \leq Y$. Thus, it is unbiased since

$$E(\hat{\theta}_2) = \frac{3}{2} \int_0^c \int_0^c \max\{x, y\} \frac{dy dx}{c^2} = 3 \int_0^c \int_0^x \frac{x dy dx}{c^2} = 3 \int_0^c \frac{x^2 dy dx}{c^2} = c.$$

The standard error of $\hat{\theta}_1$ is

$$\sigma_{\hat{\theta}_1} = \sqrt{V(X + Y)} = \sqrt{V(X) + V(Y)} = \sqrt{\frac{c^2}{12} + \frac{c^2}{12}} = \frac{c}{\sqrt{6}}.$$

The better estimator of two unbiased estimators has the smaller standard error. Thus $\hat{\theta}_2$ is better than $\hat{\theta}_1$ because

$$\sigma_{\hat{\theta}_1} = \frac{c}{\sqrt{6}} > \frac{c}{3\sqrt{2}} = \sigma_{\hat{\theta}_2}.$$

5. In a random sample of 20 Pocatello telephone customers who had trouble with their service, 12 said that the problem was fixed the same day that it was reported. Estimate the true proportion p of the customers whose problem was fixed the same day that it was reported. What is the standard error of the estimator you used? Find a 90% lower confidence bound for the true proportion p of phone problems that are fixed on the same day.

Let X be the number of respondents who reported that their phone problems were fixed on the same day. The estimate of p , the true proportion of phone problems that are fixed on the same day is

$$\hat{p} = \frac{X}{n} = \frac{12}{20} = \boxed{.6}.$$

The estimated standard error is

$$\sigma_{\hat{p}} = \sqrt{V(\hat{p})} = \sqrt{V\left(\frac{X}{n}\right)} = \sqrt{\frac{pq}{n}} \approx \hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}} = \sqrt{\frac{(.6)(.4)}{20}} = \boxed{.110}.$$

Since $n\hat{q} = 8$ is not larger than 10, we must use the unrestricted one sided CI for proportion. For $\alpha = .10$, the critical one-sided z value is $z_{\alpha} = z_{.10} = 1.282$. Thus with 90% confidence, a lower confidence bound for the true proportion p of phone problems that are fixed on the same day is given by

$$\begin{aligned} p &> \frac{\hat{p} + \frac{z_{\alpha}^2}{2n} - z_{\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha}^2}{4n^2}}}{1 + \frac{z_{\alpha}^2}{n}} \\ &= \frac{.6 + \frac{(1.282)^2}{40} - (1.282) \sqrt{\frac{(.6)(.4)}{20} + \frac{(1.282)^2}{1600}}}{1 + \frac{(1.282)^2}{20}} \\ &= \boxed{.457}. \end{aligned}$$