ECNS 309, Fall 2014

Practice Problems #2—Solutions

1. Your interpretation of a random variable will vary, but it must include some notion that a random variable is one whose outcome is unknown until an event occurs. Therefore, there is never certainty as to the future outcome, but statistical tools can be used to provide knowledge about the likelihood that a particular outcome will occur.

2. Poisson distribution

   (a) $f(5; 0.15) = \text{Prob}(x = 5) = \frac{7.5^5 e^{-7.5}}{5!} = 0.109$

   (b) $f(3; 0.25) = \text{Prob}(x = 3) = \frac{12.5^3 e^{-12.5}}{3!} = 0.0012$

3. Suppose that you trying to determine the number of different combinations of name initials.

   (a) There are 26 letters in the English alphabet. If everyone had three initials, then there are 26^3 possible unique combinations.

   (b) We already know that there are 26^3 initial combinations if people had three initials. However, if people had only two initials, then there would be 26^2 unique combinations. So, the total number of combinations if people could have two or three initials is $26^2 + 26^3$.

4. Does the conclusion follow from the premise? No. While all roses are flowers, only some flowers fade quickly. Roses may be in the set of flowers that do not fade quickly (e.g., only lillies and daffodils may fade quickly).

5. Stock portfolio.

   (a) i. $E[\text{Return}] = 6.59\%$; the central tendency / most likely return is 6.59%.

   ii. $E[\#\text{ stocks}] = 8.06$; the central tendency / most likely number of stocks in a portfolio is approximately 8.

   (b) i. $\sigma^2_{\text{Return}} = 0.00082\%^2$; the variance is $0.00082\%^2$.

   ii. $\sigma^2_{\#\text{ stocks}} = 10.38$; the variance is 10.37 stocks^2.

(c)
   i. $\sigma_{\text{Return}} = 2.86\%$; 67% of portfolio return outcomes will be within $6.59\% \pm 2.86\%$.
   ii. $\sigma_{\text{\# stocks}} = 3.22$; 67% of portfolios will have approximately 8 ± 3 stocks.

(d)
   i. $m_{\text{Return}} = \frac{\sum_{i=1}^{11}(Return_i - E[\text{Return})^3 \cdot Prob[Return_i])}{\sigma^3_{\text{Return}}} = 0.502$; the Returns pdf is positively skewed, implying that most returns will be relatively low, but some may be high. Positive skeweness implies that the mean is greater than the median.
   ii. $m_{\text{\# stocks}} = \frac{\sum_{i=1}^{11}(\#stocks_i - E[\text{\# stocks})^3 \cdot Prob[\#stocks_i])}{\sigma^3_{\text{\# stocks}}} = 0.037$; the number of stocks pdf is positively skewed, implying that most returns will be relatively low, but some may be high.

6. Analysis of sales and prices.
   (a) $\overline{\text{Sales}} = 183$; $\overline{\text{Price}} = \$23218$

   (b) $s^2_{\text{Sales}} = 2899$; $s^2_{\text{Price}} = 22410114$

   (c) $s_{\text{Sales}} = 54$; $s_{\text{Price}} = \$4734$

   (d) $CV_{\text{Sales}} = 0.29$; $CV_{\text{Price}} = 0.20$

   (e)
      i. $H_0: \overline{\text{Sales}} - 210 = 0$

      ii. $H_a: \overline{\text{Sales}} > 210$

      iii. $t_{\text{stat}} = \frac{183-210}{54/\sqrt{14}} = -1.871$

      iv. $t_{\text{crit, 1-tailed, df=13, }\alpha=0.05} = 1.771$

      v. The t_{stat} is negative, so it is clearly not greater than the critical value. Therefore, we cannot reject the null hypothesis in favor of the alternative hypothesis. That is, there is not enough statistical evidence to conclude with a 95% confidence level that average sales in Bozeman are greater than 210.

(f)
   i. $H_0: \ \overline{\text{Price}} - 21000 = 0$

   ii. $H_a: \ \overline{\text{Price}} \neq 21000$

   iii. $t_{\text{stat}} = \frac{23218-21000}{4733/\sqrt{14}} = 1.70$

   iv. $t_{\text{crit, 2-tailed, df=13, }\alpha=0.05} = 2.160$

   v. The $|t_{\text{stat}}| < 2.160$, so we cannot reject the null hypothesis in favor of the alternative hypothesis. That is, there is not enough statistical evidence to conclude with a 95% confidence level that average prices in Bozeman are not equal to \$21,000. We would infer that the slightly higher sample average price we found, \$23,218, was likely higher due to a statistical anomaly.

(g) $S_{\text{Sales, Price}} = -208995$

(h)
   i. $\rho_{\text{Sales, Price}} = -0.82$

   ii. $H_0: \ \rho_{\text{Sales, Price}} - 0 = 0$

   iii. $H_a: \ \rho_{\text{Sales, Price}} \neq 0$

   iv. $t_{\text{stat}} = -0.82\sqrt{\frac{14-2}{1-(-0.82^2)}} = -4.96$

   v. $t_{\text{crit, 2-tailed, df=12, }\alpha=0.05} = 2.179$

   vi. The $|t_{\text{stat}}| > 2.160$, we reject the null hypothesis in favor of the alternative hypothesis. That is, there is enough statistical evidence to suggest that the correlation between sales and prices is statistically different than zero at a 95% confidence level.

(i)
   i. $\hat{\beta} = \frac{S_{\text{Sales, Price}}}{s^2_{\text{Price}}} = -0.009$

ii. $\hat{\alpha} = \overline{\text{Sales}} - \hat{\beta}\overline{\text{Price}} = 384.8$

(j) The estimate $\hat{\beta} = -0.009$ suggests that, on average, a \$1 increase in sales price is correlated with a 0.009 reduction in sales. We can scale things to get a more realistic estimate. That is, on average, a \$1,000 increase in sales price is correlated with a 9 car reduction in demand.

7. Scaling covariance and correlation.

(a) $S_{y,x} = S_{y,y} = \frac{1}{n-1}\sum(y_i - \bar{y})^2 = s_y^2$. The covariance will equal the variance.

(b) $\rho_{y,x} = \rho_{y,y} = \frac{s_y^2}{s_y s_y} = 1$. As we would expect, when $y = x$, the correlation will be perfect.

(c) $S_{y,x} = S_{y,\alpha y} = \alpha s_y^2$;

$\rho_{y,\alpha y} = \frac{\alpha s_y^2}{s_y \alpha s_y} = 1$. The correlation between any variable and a scaled form of that variable will be perfect.

8. Empirical analysis

(a) Since I created the dataset, I didn't need to download it.

(b) and (c)

Number of coffeeshops		Median Income (in \$1,000)	
Count	49.00	Count	49.00
Mean	19.73	Mean	30.55
Median	5.00	Median	30.17
Variance	1287.16	Variance	13.85
Standard Deviation	35.88	Standard Deviation	3.72
Coefficient of Variation	1.82	Coefficient of Variation	0.12

Covariance = 68.65
Correlation = 0.51
Correlation $t_{\text{stat}} = 4.11$
Correlation $t_{\text{crit}} = 2.021$
Reject null that correlation is not statistically different from zero

(d) Linear function: Coffeeshops $= \alpha + \beta$Median Income $+ e$

(e) Regression results in standard presentation form (standard errors in parentheses):

$$\begin{array}{lcccl} \text{Coffeeshops} & = & \text{-131.67} & + & 4.96 \quad \text{Median Income} \\ & & (37.11) & & (1.21) \end{array}$$

(f) Testing statistical significance $H_0: \ \hat{\alpha} - 0 = 0$

$H_a: \ \hat{\alpha} \neq 0$

$t_{\text{stat}} = \frac{-131.67-0}{37.11} = -3.55$

$t_{\text{crit, 2-tailed, df=48, }\alpha=0.05} = 2.021$

The $|t_{\text{stat}}| > 2.021$, we reject the null hypothesis in favor of the alternative hypothesis. That is, there is enough statistical evidence to suggest that the intercept coefficient is statistically different than zero at a 95% confidence level.

$H_0: \ \hat{\beta} - 0 = 0$

$H_a: \ \hat{\beta} \neq 0$

$t_{\text{stat}} = \frac{4.96-0}{1.21} = 4.11$

$t_{\text{crit, 2-tailed, df=48, }\alpha=0.05} = 2.021$

The $|t_{\text{stat}}| > 2.021$, we reject the null hypothesis in favor of the alternative hypothesis. That is, there is enough statistical evidence to suggest that the slope coefficient is statistically different than zero at a 95% confidence level.

9. <u>On average</u>, a $1,000 increase in median income is positively correlated with approximately a 5 coffeeshop increase in a county.

10. $\varepsilon_{\text{Coffeeshop,Income}} = \frac{dC}{dI}\frac{\bar{I}}{\bar{C}} = 4.96\left(\frac{30.55}{19.73}\right) = 7.68\%$. This implies that coffeeshop demand is quite substantially income elastic. For a 1% change in income, there is a 7.68% change in the number of coffeeshops, suggesting that small changes in income would lead to significant increases in take-out coffee demand. However, small reductions in income (perhaps in a recession) would lead individuals to significantly reduce their take-out coffee consumption.

11. Given the significantly high elasticity value, one potential inference is to build a coffeeshop in a county where there is the fastest growth in income.