Key to Sample Stat 224 Test II

Prof. Callan

1. a) First, compute the cumulative probabilities $P(X \leq k)$. This determines the range to simulate k.

k	$P(X \leq k)$	range to simulate k
0	.12	$[00, 12)$
1	.24	$[12, 24)$
2	.84	$[24, 84)$
3	1.00	$[84, 100)$

The simulated values of X are thus 1, 3, 1, 2, 2, 2, 1, 1, 3, 2 and there are 4 twos in this list, a little below the mathematical expectation, which is $np = 10 \times .6 = 6$.

1. b) (i) $E(3X - 2Y + Z) = 3E(X) - 2E(Y) + E(Z) = 6.0 - 2.4 + 1.6 = 5.2$
(ii) $Var(3X - 2Y + Z) = 9Var(X) + 4Var(Y) + Var(Z) = 9(.64) + 4(.04) + .16 = 6.08$.

$3X - 2Y + Z$ has a normal distribution with mean $\mu = 5.2$ and standard deviation $\sigma = \sqrt{6.08} = 2.5$, (rounded off to one decimal place, since the data were given to one decimal place).

2. a) For $0 < x < 1$, the marginal probability density function for X is

$$f_1(x) = \int_{y=0}^{y=2x} \frac{3}{4}(x+y)\,dy = \frac{3}{4}(xy + \frac{y^2}{2})\bigg|_{y=0}^{y=2x} = 3x^2$$

b) $P(Z < a) = .8664 + (1 - .8664)/2 = .9332$. From the Normal Tables, $a = 1.50$.

3. a) Work in units of 1000 miles throughout. The hypotheses to be tested are

$$\begin{aligned} H_0 &: \mu = 40 \\ H_1 &: \mu > 40 \end{aligned}$$

The data yield sample mean $\overline{x} = 40.70$ and sample standard deviation $s = 2.48$ (after entering data in your calculator), and the test statistic is

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{40.7 - 40}{2.48/\sqrt{12}} = .977$$

The cutoff value is $t_{.05}$ at $\nu = n - 1 = 11$ degrees of freedom, which is 1.796 from Table 4.

Is .977 > 1.796? No. Conclusion: accept H_0 and reject the manufacturer's claim.

3. b) (i) Yes, because .042 is < .05.

(ii) 2(.042)=.084 (diluted significance when altrnative is two-sided)

(iii) small (the smaller the P-value of the sample, the stronger the evidence against the null hypothesis)

4. a) The hypotheses to be tested are

$$\begin{aligned} H_0 &: \sigma_1 = \sigma_2 \\ H_1 &: \sigma_1 \neq \sigma_2 \end{aligned}$$

The test statistic is

$$F = S^2_{\text{big}}/S^2_{\text{small}} = .28/.14 = 2$$

and we reject H_0 if $F > F_{\alpha/2} = F_{.05}$ at $\nu_{\text{numerator}} = 11$ and $\nu_{\text{denominator}} = 9$, which is 3.10. Since $F \not> 3.10$, we accept H_0 and conclude the variance of the yield is not affected by the catalyst.

b) Here X is the number of rolls till the first six appears, $p = 1/6$, $E(X) = 6, \text{Var}(X) = 30$. So $\overline{X}$ has mean 6, standard deviation $\sqrt{30/100} = .548$ and, by CLT, $\overline{X}$ is approx. normal. So

$$P(\overline{X} > 6.5) = P\left(\frac{\overline{X} - 6}{.548} > \frac{6.5 - 6}{.548}\right) = P(Z > .91) = 1 - .8186 = .1814$$

5. We need $n \geq (z_{\alpha/2}/E)^2 p(1-p)$. Knowing nothing about p, we cater to the largest $p(1-p)$ can be, namely 1/4 (when $p = 1/2$) and use $n = (1.96/.02)^2(1/4) = 2401$.

(Note: for $E = .03$, often used in practice, it turns out $n = 1068$)

(i) The parameter p is not a random variable—it either lies in the reported interval or it does not. So it is incorrect to speak of *probabilities* concerning p.

(ii) The interval $\hat{p} \pm z_{.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ (its endpoints are random variables) is a random interval that contains (or covers) p with probability .95. If many such intervals were constructed (say by different polling organizations), on average 95 out of 100 of them *would* cover p.