1.

Applied Probability 2 Exam Paper 2011

Q 4 A. $u = 10.5 \quad \sigma^2 = 1.09$

$u \pm t_{(1-\alpha/2, 7)} \frac{s}{\sqrt{n}}$

$10.5 \pm 3.499 \frac{\sqrt{1.09}}{\sqrt{8}}$

$(9.208, 11.792)$ 99% CI

b $\frac{(n-1)s^2}{\chi_{(0.025, 7)}}, \frac{(n-1)s^2}{\chi_{(0.975, 7)}}$

$\frac{7(1.09)}{16.013}, \frac{7(1.09)}{1.690}$

$(0.476, 4.513)$ 95% CI for variance

C For a:

If we repeat the above tests in the similar manner, 99% of the time we are confident that our interval contains the true mean growth rate of plant

For b:

95% of time when we simulate under similar conditions we are confident that true variance lies within this interval

D Bootstrap samples are generated from the existing data by resampling from that data to create a new sample

Can be done in excel by using index function

index(range values in, Sample Size*RAND()).

2

4 E Digit 4 5 5 1 3 6 4 4
Value: 12.0, 8.9, 8.9, 10.6, 11.4, 11.1, 12.0, 12.0
This is our new sample

F - Generate a large number of these bootstrap samples thus
- Calculate mean for each sample
- Rank the means in ascending order
- 95% CI is created from (2.5th percentile, 97.5th percentile)
thus 95% of data lies within this interval

There are only 8 values in the sample thus the central limit may not hold accurately

With repetition of bootstrap it is likely to get more extreme values thus the interval could be wider

Q5 A Only two outcomes: Success or failure
N - independent trials
Same probability of success in each trial
This describes binomial distribution which is applicable in this situation

b P_1 = first group not brown $\frac{25}{40} = 5/8$
$P_2 = \frac{30}{60} = 1/2$

C $H_0: P_1 = 0.5$ Vs $H_1: P_1 > 0.5$
$P_1 = 5/8$
$$z = \frac{5/8 - 0.5}{\sqrt{\frac{5/8(1-5/8)}{40}}} \quad \frac{1/8}{\frac{\sqrt{6}}{32}} = 1.633$$

$t_{critical} = (0.95, 39) = 1.684$
one side

$t_{calculated} < t_{critical}$, no evidence against H_0

3

Applied Probability 2 Exam Paper 2011

Q5 d i. If $p_1 = p_2$ then both samples come from the same distribution, and hence p_1 should equal p_2.

Then p would be $\frac{25+30}{40+60} = 0.55$ which would be a reasonable allow

ii. - Under assumption both come from same distribution
- Thus bootstrap samples can be generated from the combined 100 values
- Done using excel's index function (range, sample size * RAND())
- Generate a bootstrap of 40 for hypothesis and 60 for non all from the combined sample

iii. 95% CI for bootstrap is (~~2.5th percentile, 97.5th percentile~~) dependent on 95th percentile
= 0.160

zero is less than this 0.160 so we cannot reject H_0 that $P_1 = P_2$.

e Type 1: probability of rejecting a null hypothesis when it is true

Type 2: probability of accepting an alternate hypothesis when it is false

Type I error can be reduced by ~~decreasing alpha, the~~ decreasing alpha, has effect of increasing probability of making a type II error.

4

6a i. We have data $x_1, x_2 \dots x_n$ normally distributed

mean $\mu = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\sigma^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

From properties of normal distribution, the values:
$z_1 = \frac{x_1 - \mu}{\sigma}$, $z_2 = \frac{x_2 - \mu}{\sigma}$, ..., $z_n = \frac{x_n - \mu}{\sigma}$ are normally distributed with mean 0, variance
z-score

In practice, with a data set of size n, it is easiest to look at the $100\frac{1}{n+1}$th, $100\frac{2}{n+1}$th, ... $100\frac{n}{n+1}$th percentiles

because they correspond exactly to each data point in order.

If we do a scatter plot of the z-score of the k^{th} ordered observation in order against the expected z-score of $k/n+1$, all the above implies that they should lie on straight line $y = x$

ii. The data does not lie on a straight line, thus it can be concluded that the data is not normally distributed

iii. If we take the log of the values, it is likely the values will then be normally distributed
We could also: -square root, square or other power
or use inverse $1/x$ etc

Applied Probability 1 Exam Paper 2011

Q6 B i. X value: 1 | 2 | 3 | 4

$p(x) =$ 0.2 | 0.18 | 0.22 | 0.4

y value: 1 | 2

$p(y)$ 0.85 | 0.15.

ii. $Cov = E(xy) - E[x]E[y]$

$E[x] = 1(0.2) + 2(0.18) + 3(0.22) + 4(0.4) = 2.82$

$E[y] = 1(0.85) + 2(0.15) \quad 1.15.$

$E[xy] = 1(0.1)(1) + 1(0.15)(2) + 1(0.2)(3) + 1(0.4)4$
$+ 2(0.1)(1) + 2(0.03)(2) + 2(0.02)(3) + 2(0)4 = 3.04$

$3.04 - 2.82(1.15) = -.203$

iii. $E[x^2] = 9.3 \quad var = 9.3 - 2.82^2 = 1.3476 = var[x]$

$E[y^2] = 1.45 \quad var = 1.45 - 1.15^2 = 0.1275.$

correlation $= \dfrac{-0.203}{\sqrt{0.1275(1.3476)}} = -0.484 \Rightarrow$ relationship exist not indepen

$E(Y=1, X=1) = 0.1$

$E(Y=1)E(X=1) = 0.85(0.2) = 0.17 \neq 0.1$ NOT indeped

iv. 1,1 1,2 1,3 1,4 2,1 2,2 2,3 2,4

✓ ✓ ✓ ✓ ✓

$0.1 + 0.15 \quad 0.20 + 0.1 + 0.03$

$p(x+y \leq 4) = 0.59$

C i. $X \sim N(3, 2^2)$

ii. $\dfrac{0-4}{1} \leq y \leq \dfrac{3-4}{1} \quad -4 \leq z \leq -1$

$(1 - 0.8413) - 1(0)$

$= \quad 0.1387 = p(0 \leq y \leq 3)$

6.

iii. $\bar{U} = 4 + 0.8\left(\frac{1}{2}\right)(2-3)$

$4 - \frac{2}{5} = 3.6$

$\sigma^2 = (1-p^2)\sigma_y^2$

$(1-0.8)1 = 0.2$

$\sim N(3.6, (\sqrt{0.2})^2)$

$P(0 \le y \le 3 \mid x=2) =$

$\frac{0-3.6}{0.2} \le z \le \frac{3-3.6}{0.2}$

$-1.8 \le z \le -0.3$

$(1-0.6179) - (1-0.7881)$

$0.3821 - 0.2119 = 0.1802 = p(0 \le y \le 3 \mid x=2$