27/3/14. Applied Probability 2 Exam paper 2013

Q1 $n = 10$
$\bar{x} = 267$
$s^2 = 218.7$

a CI for mean. $\bar{x} \pm t_{0.005} \frac{s}{\sqrt{n}}$
99% CI

$267 \pm 3.24 \times \frac{\sqrt{218.7}}{\sqrt{10}}$
267 ± 15.2

b 99% CI
If I repeated this experiment many times, each time I would most likely observe different score and so compute a different 99% CI. But in 99% of the intervals, the true mean score would lie in it.

b 3 out of 10 games proportion is 3/10.

c From lab 1
INDEX(A1:A10, Ceiling (10*RAND()))
↑ int from 1 to 10.

Generates 1 random observation
Reapeat 10 times

D For each bootstrap sample, calculate proportion of values that are >40
This gives us 1000 proportions

A 95% interval is got by taking 2.5% = 25th smallest and 2.5% = 25th largest (2.5 and 97.5) e $\hat{p} = 0.3 \quad n = 10$

$t_{0.025, n-1} = 2.262$

$0.3 \pm 2.262 \times \sqrt{\frac{0.3(0.7)}{10}}$

0.3 ± 0.328

$(-0.028, 0.628)$

F. Bootstrap may be better here:

- Formula based on assuming $\hat{p}$ is normally distributed by central limit theorem. But we have a small sample (10) so it may not be applicable

- Lower ~~interval~~ limit of interval from formula is negative, which is impossible for a proportion

Q2 A-B testing - web page.

A Two Sample t-Test.

$H_0 \quad \mu_1 = \mu_2$

$H_1 \quad \mu_1 \neq \mu_2$ (two-tailed test)

1st Sam $\bar{x} = 45.5 \quad S_1 = 12 \quad n = 50$

2nd Sam $\bar{y} = 48.7 \quad S_2 = 10 \quad m = 40$

$S^2_{combined} = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} \quad \frac{49(12^2) + 49(10^2)}{88} = (11.16)^2$

$S = \sqrt{S^2} = 11.86$

$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n} + \frac{1}{m}}} \quad \frac{45.5 - 48.7}{11.16\left(\sqrt{\frac{1}{50} + \frac{1}{40}}\right)} = -1.35$

5% level significance

Reject H_0 if $|t| > t_{0.025, 88}$

$|t| \not> = 1.984$

Do not reject H_0.

27/3/14. Applied Probability Exam Paper 3

Q 2b In notes

This is the probability of observing $t = -1.35$ or a value ~~more~~ more extreme. e.g. stronger evidence against Ho. In this case where Ho is true

Since this is a 2 tailed test, this mean $t < -1.35$ ~~or t > 1.35~~ or $t > 1.35$

c Did something similar in 2^{nd} lab.

i. If Ho true, then data in both samples is coming from the same population, ie it can be merged

ii. If ~~Ho true~~ look at 2.5 and 97.5^{th} percentile values then there is a 5% probability of lying outside these values

Our observed difference is $45.5 - 48.7 = -3.5$, -3.5 does not lie outside 2.5th 97.5th percent value range of $(-3.94, 4.57)$
$\Rightarrow$ fail to reject Ho

Di 2 Sample t-Test not appropriate as it assumes all customers have same average spending with design 1; and similarly with design 2
Also data are "paired", across designs by customers $\Rightarrow$ sensible to use paired t-test.

ii Ho: $\mu_{1i} = \mu_{2i}$ for $i = 1 \ldots 20$ — μ_{1i} average spent by customer i on design 1, μ_{2i} average spent by customer i on design 2

H_1: $\mu_{1i} \neq \mu_{2i}$

4

$t = \dfrac{d}{s_d/\sqrt{n}} = \dfrac{-2}{5/\sqrt{20}} = -1.79$

5% level signif

Reject H_0 if $|t| > t_{0.025,19} = 2.093$

$\Rightarrow$ do not reject H_0

1/4/14. Applied Probability 2013 exam Q3

Q3 $\text{Prob}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$

Likelihood = Probability of the observation = $\text{Prob}(k_1, k_2, \ldots, k_{100})$

$= \text{Prob}(k_1) \times P(k_2) \times \ldots \times P(k_{100})$

Always assumes independence

$\frac{\lambda^{k_1} e^{-\lambda}}{k_1!} \; \frac{\lambda^{k_2} e^{-\lambda}}{k_2!} \; \ldots \ldots \; \frac{\lambda^{k_{100}} e^{-\lambda}}{k_{100}!}$

$= \frac{\lambda^{k_1+k_2+\ldots+k_{100}} e^{-100\lambda}}{k_1! \times k_2! \times \ldots \times k_{100}!}$

ii log likelihood

$\text{Log} \frac{\lambda^{k_1+k_2+\ldots+k_{100}} e^{-100\lambda}}{k_1! \times k_2! \times \ldots \times k_{100}!}$

$= \text{Log}(\lambda^{k_1+k_2+\ldots k_{100}}) + \log(e^{-100\lambda}) - \log(k_1! \times k_2! \times k_{100}!)$

$(k_1+k_2+\ldots k_{100}) \log(\lambda) \quad + \quad -100\lambda \quad - \log(k_1! \times k_2! \times \ldots \times k_{100}!)$

$\sum_{i=1}^{100} k_i \log(\lambda) \quad -100\lambda \quad - \log\left(\prod_{i=1}^{100} k_i!\right)$

iii $\hat{\lambda}$ is value of λ that maximises log likelihood (or likelihood).

$\frac{dL}{d\lambda} = \frac{\sum_{i=1}^{100} k_i}{\lambda} - 100 - 0$

Solving $\frac{dL}{d\lambda} = 0 \rightarrow \lambda = \frac{1}{100} \sum_{i=1}^{100} k_i$

second derivative $\frac{-1}{\lambda^2} \sum_{i=1}^{100} k_i \rightarrow$ negative $\rightarrow$ maximum

$\uparrow \geq 0 \rightarrow$ positive

6.

3 b i. Bookwork - check if normally distributed

ii. points lie on straight line.
That data and q-q plot are consistent with normally distributed observations

C i. $p(x)$: Distribution of X is the column sum

$P(x=2) = 0.1$
$P(x=4) = 0.4$
$P(x=5) = 0.5$

$P(y)$ Distribution of y is row sum (marginal)
$p(y=-1) = 0.15$ $p(y=0) = 0.45$ $p(y=4) = 0.4$

ii Conditional dist $X \mid Y=0$ $= \frac{P(x, y=0)}{p(y=0)}$

$\frac{P(x, y=0)}{0.45}$ $P(x=2 \mid y=0) = \frac{P(x=2, y=0)}{0.45} \quad \frac{0.1}{0.45} = 2/9$

$P(x=4 \mid y=0) = \frac{P(x=4, y=0)}{0.45} \quad \frac{0.25}{0.45} = 5/9$

$P(x=5 \mid y=0) \quad 0.1/0.45 = 2/9$

iii Correlation of X and Y $= \frac{\text{covariance}(x,y)}{\sqrt{\text{var}(x)\,\text{var}(y)}}$

$E(X) = (2 \times 0.1) + (4 \times 0.4) + 5(0.5) = 4.3$
$E(y) = 1.45$
$E(xy) = \sum_{\text{all } x} \sum_{\text{all } y} xy\, P(x,y) \quad -2(0) + 4(-1)(0.05) + 5(-1)(0.1) \ldots$ all y $= 6.9$
$E(x^2) = (4 \times 0.1) + (16 \times 0.4) + (25 \times 0.5) = 19.3$
$E(y^2) = 6.55$

$p(x,y) = \frac{E(xy) - E(x)E(y)}{\sqrt{(E(x^2) - E(x)^2)(E(y^2) - E(y)^2)}}$

$\frac{6.9 - (4.3 \times 1.45)}{\sqrt{(19.3 - 4.3)^2 \; (6.55 - 1.45)^2}} = 0.35$ between ± 1

7 Applied Probability

Q3C(ii) Show x and y not independent

- covariance is <u>not</u> zero $6.9 - (4.3)(1.45) = 0.665$.
if independent then $cov(x,y) = 0$
$cov \neq 0 \Rightarrow$ not independent

Directly show that $p(x,y) \neq p(x)p(y)$

eg. $p(x=2, y=-1) = 0$

$p(x=2) \times p(y=-1)$ $0.1 \times 0.15 = 0.015 \neq 0$ not independent

only have to repeat this until you find <u>not independent</u> value

either way is ok