Stat 13
Homework 7

7_1)

a) This is NOT representative of New Zealander computer usage. The random sample was only of subscribers to Consumer Magazine so it's only applicable to that population.

b) $300/2730 = 0.1099$

c) $109/2730 = 0.0399$

d) P(drawing) = 0.1099
P(publishing) = 246/2730 = 0.0901

H_0: P(d) - P(p) = 0
H_1: Not equal

Use z-score since it is proportions. SE given by one sample many responses.

$z_0 = (0.1099 - 0.0901 - 0)/\sqrt{[(0.1099+0.0901-(0.1099-0.0901)^2)/n]}$
$= 2.3156$
$P(Z>|z_0|) = 0.02$ = p-value

95% CI: $(0.1099 - 0.0901) \pm Z(0.95) * \sqrt{[(0.1099+0.0901-(0.1099-0.0901)^2)/n]}$
$= [0.0030,0.0366]$

P-value says that if the null hypothesis was true, then a result this extreme or more extreme would only happen one time out of fifty, so it's pretty rare. We can reject the null hypothesis at the 5% level. The 95% CI doesn't contain zero so it appears there is actually a difference, in line with our p-value.

7_2)

a) Theta: P(white) - P(gypsy)

b) H_0: theta = 0
H_1: theta not equal to zero

c) $496/886 - 74/452 = 0.56 - 0.49 = 0.07$

d) Z-score since proportions, SE given by independent samples.

$z_0 = (0.07 - 0)/\sqrt{(0.56(1-0.56)/886 + 0.49(1-0.49)/152)}$

= 0.07/0.043 = 1.63

p-value = P(Z>z_0) = P(Z>|1.63|) = 2*(1-0.948) = 2*0.052 = 0.104

e) If we took many samples of this size then 10.4% of them would give a result this extreme or more extreme for the difference in proportions assuming the null hypothesis is true (the difference is zero).

f) There is weak evidence for the difference but not enough for the 95% level, so we would conclude that there is no difference in the proportions.

g) 95% CI: 0.07 +/- 1.96*(0.043) = [-0.0143,0.1543]

h) If many samples are taken then 95% of the CI would capture the true mean, so with 0.95 probability the true mean lies in the interval [-0.000735,0.1407]. Since this interval contains zero, we cannot reject the hypothesis of no difference.

7_3) rice ~ N(mu,2.7)
mu set at 506
Want packets in interval [500,512] 95% of time.

a) P(500<X<512) = P[(500-506)/2.7<Z<(512-506)/2.7] =
1-P(Z<|6/2.7|) = 1-2*P(Z<-2.22) = 1-2*0.013 = 1-0.026 = 0.974>0.95
so the manager's needs will be met.

b) i) X~N(506,2.7/sqrt(25)) = N(506,2.7/5)
ii) No, CLT was not needed since we know the sum of Normal variables is normally distributed.
iii) 95% CI: [506-1.96*2.7/5,506+1.96*2.7/5] = [504.94,507.06]

c) i) P_hat~N(0.05,sqrt[0.05(1-0.05)/800])=N(0.05,0.0077)
ii) Yes, needed the CLT to approximate the binomial distribution by the normal.
iii) 95% CI: [0.05+1.96*0.0077,0.05-1.96*0.0077]
= [0.0349,0.0651]

d) i) Output from STATA
plot in units of .1

```
49** | 83,92
50** | 08,14,16,19
50** | 24,26,28,29,29,34,39,39
50** | 40,44,45,54,55
50** | 63,76
50** |
51** | 03,04,19
```

51** |
51** | 50

ii) No, it doesn't appear to be symmetrical, it appears skewed to the right.

iii) Gaps and skewness to the right.

iv) No it isn't likely.

e) X_bar ~ N(506,0.54)
H_0: mu = 506
H_1: Not equal

z_0 = (504.5-506)/0.54 = -2.78

p-value = P(Z>|z_0|) = 2*(1-0.997) = 0.006

We can definitely reject the null hypothesis at the 99% level, the mean output from the machine is not 506.

f) p-hat = 26/800 = 0.0325
H_0: p = 0.05
H_1: p not equla to 0.05

z_0 = (0.0325-0.05)/0.0077 = -2.27

p-value = P(Z>|z_0|) = 2*(0.012) = 0.024

We can reject the null hypothesis at the 95% level, the mean output from the machine is not 506.

g) This should just summarize the p-values from the proportion and mean, stating what they mean and that it is not in line with the null hypotheses.

7_4)
a) i) 45-54
ii) 35-44: max over age groups[count(attended 1-3 years) +
count(attended 4+ years)]

b) i) (did not complete HS + Completed HS)/ Total = (58174+29539)/172214
= 0.5093
ii) (Completed HS + (35-44) - (completed HS and (35-44))/Total =
(58174+44460-15136)/172214 = 0.5081

c) (25 to 34 + 35 to 44 + 45 to 54 of completed HS)/(total completed HS) =

(12569 + 15136 + 10943)/58174 = 0.6643

d) (Didn't complete HS + Completed HS of 25 to 34)/(total 25 to 34) =
(4754 + 12569)/39354 = 0.4402

e) P(25 to 34 w/ 4+ yrs Univ.) = P_u = 2444/39354 = 0.0621
P(64+ w/ did not complete HS) = P_hs = 10580/32086 = 0.3297

Z-score Independent samples for SE.

H_0: P_hs - P_u = 0
H_1: Not equal to zero

SE = sqrt[0.0621(1-0.0621)/39354 + 0.3297(1-0.3297)/32086]
 = 0.00289

z_0 = [(0.3297-0.0621) - 0]/0.00289 = 92.59

P(Z>z_0) = 0

Definitely statistically significant, this would never happen (vanishingly small probability) under the null hypothesis of equality.