Lab 09 - Correlations by Hand

(1)

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$	y	$\bar{y}$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
587000	666100	-79100	6256810000	148779	209408.333	-60629.333	3675916060	4795780267
706600	666100	40500	1640250000	265632	209408.333	56223.667	3161100693	2277058500
625300	666100	-40800	1664640000	203085	209408.333	-6323.333	39984544.4	257992000
680000	666100	13900	193210000	225056	209408.333	15647.667	244849473.2	217502567
634000	666100	-32100	1030410000	191423	209408.333	-17985.333	323472213.9	577329199
700000	666100	33900	1149210000	221912	209408.333	12503.667	156341680.9	423874301
722000	666100	55900	3124810000	222431	209408.333	13022.667	169589848	727967069
657000	666100	-9100	82810000	195999	209408.333	-13409.333	179810219.6	122024933
683000	666100	16900	285610000	210358	209408.333	949.667	901866.8	16049367.2
$\Sigma x = 5994900$				$\Sigma y = 1884675$				$\Sigma = 9415578203$
$\bar{x} = 666100$				$\bar{y} = 209408.333$				

there may be rounding error here

$\Sigma(x-\bar{x})^2 = 15427760000$

$s^2 = 1928470000$

$s_x = 43914.34845$

$\Sigma(y-\bar{y})^2 = 7951966601$

$s^2 = 993995825.1$

$s_y = 31527.69933$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y} = \frac{9415578203}{(8)(43914.34845)(31527.69933)} = \frac{9415578203}{11076146995} = .850$$

$r^2 = .723$

$v = 9-2 = 7$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{v}}} = \frac{.85}{\sqrt{\frac{1-.723}{7}}} = \frac{.85}{\sqrt{\frac{.277}{7}}} = \frac{.85}{\sqrt{.040}} = \frac{.85}{.2} = 4.25$$

There is a strong, positive correlation ($r=.85$, $p=.004$) between population and turnout. Population explains 72.3% of the variation in turnout. Larger districts have larger numbers of voters who show up to the polls.

②

x	$\bar{x}$	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
1	.333	.667	.444889	-60629.333	-40439.765
0	.333	-.333	.110889	-56223.667	-18722.481
1	.333	.667	.444889	-6323.333	-4217.663
0	.333	-.333	.110889	15647.667	-5210.673
1	.333	.667	.444889	-17985.333	-11996.217
0	.333	-.333	.110889	12503.667	-4163.721
0	.333	-.333	.110889	13022.667	-4336.548
0	.333	-.333	.110889	-13404.333	4465.308
0	.333	-.333	.110889	949.667	-316.239

$n = 3$ $\quad \Sigma = 2.000001$ $\quad \Sigma = -84938$

$\bar{x} = .333$ $\quad s^2 = .25000013$

$s = .500$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y} = \frac{-84938}{(8)(.5)(31527.69933)} = \frac{-84938}{126110.797} = -.674$$

$r^2 = .432$

$v = 9-2 = 7$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{v}}} = \frac{-.674}{\sqrt{\frac{1-.432}{7}}} = \frac{-.674}{\sqrt{\frac{.568}{7}}} = \frac{-.674}{.08114286} = -8.306$$

There is a strong negative correlation ($r = -.674$, $p < .001$) between party and turnout. Party accounts for 43.2% of the variation in turnout. Democratic races in 2012 had lower absolute turnout than those when Republicans won.

(3) There is a positive relationship between price and engine size - cars with larger engines are more expensive. This is a positive association, and may be moderate. The trend line is a decent but not perfect fit for the data.

(4) There is a negative relationship between price and mileage - to some extent, cheaper cars have better fuel efficiency. However, the trend line is not a particularly good fit for the data. The relationship is likely moderate.