POLS 6481. Research Design and Quantitative Methods II

Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups

Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

Outline

1. A Single Dummy in a Regression
2. Multiple Dummies for a Single Dimension in a Regression
3. Hypothesis Testing for Unidimensional Dummies
4. Example: *'Looking Good on Course Evaluations' part 2*
5. Review of Poisson Probability Distribution & Poisson Regression
6. Example: *Economic Model of Crime*

POLS 6481. Research Design and Quantitative Methods II

Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups

Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

1. A Single Dummy in a Regression

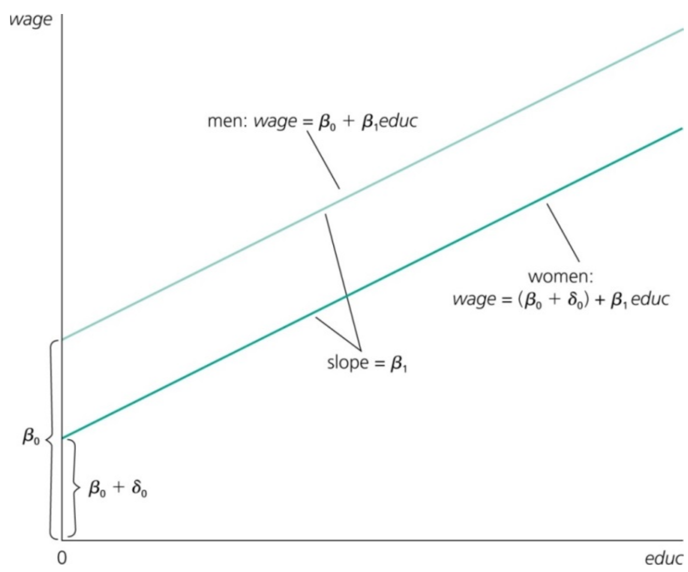
Combined model: $wage = \beta_0 + \delta_0 \cdot female + \beta_1 educ + u$

men only: $wage = \beta_0 + \beta_1 educ + u$

women only: $wage = (\beta_0 + \delta_0) + \beta_1 educ + u$

Intercept shift interpretation: $E(wage \mid female{=}1, educ = 0) = \beta_0 + \delta_0$

$E(wage \mid female{=}0, educ = 0) = \beta_0$



POLS 6481. Research Design and Quantitative Methods II

Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups

Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

1. A Single Dummy in a Regression

Combined model: $wage = \beta_0 + \delta_0 \cdot female + \beta_1 educ + u$

men only: $wage = \beta_0 + \beta_1 educ + u$

women only: $wage = (\beta_0 + \delta_0) + \beta_1 educ + u$

Treatment effect interpretation: $\delta_0 = \mathrm{E}(wage \mid female{=}1, educ) - \mathrm{E}(wage \mid female{=}0, educ)$

POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

1. A Single Dummy in a Regression

Example: *Gender and Wages* (Wooldridge Example 7.1; use WAGE1.DTA)

Practical effect of adding control variables:

$$\begin{aligned}\widehat{wage} &= \underset{(.21)}{7.10} - \underset{(.30)}{2.51}\, female \\ n &= 526,\ R^2 = .116.\end{aligned} \qquad [7.5]$$

$$\begin{aligned}\widehat{wage} &= \underset{(.72)}{-1.57} - \underset{(.26)}{1.81}\, female + \underset{(.049)}{.572}\, educ + \underset{(.012)}{0.25}\, exper + \underset{(.021)}{.141}\, tenure \\ n &= 526,\ R^2 = .364.\end{aligned} \qquad [7.4]$$

Adding education, experience and tenure helps avoid omitted variable bias – if women have lower values of these variables than men, on average. (Look back at equations for OVB; if *female* is negatively correlated to *educ*, *exper*, *tenure*, then …)

Adding education, experience and tenure improves model fit and reduces standard errors, even after accounting for how correlation among indep. variables inflates standard errors.

POLS 6481. Research Design and Quantitative Methods II

Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups

Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

1. A Single Dummy in a Regression – some comments

Note 1: 👍 regression analysis replaces ANOVA and ANCOVA for modeling control group versus experimental/treatment group(s)

Note 2: 👌 changing base/benchmark group (sign of *t* reverses)

Note 3: 👎 a 'dummy' variable with no theoretical meaning… e.g., American 'exceptionalism'

Note 4: ☠ Caution about endogenous treatment selection… e.g., challenger 'quality'

POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

2. Multiple Dummies for a Single Dimension in a Regression

Consider an *ordinal* explanatory variable…

e.g., cities' and local governments' credit ratings (0 = worst, …, 4 = best) [p. 237–238, 5th ed.]

Option A. regress municipal bond ratings on a single variable ($CR \in \{0, 1, 2, 3, 4\}$)…

$$MBR = \beta_0 + \beta_1 CR + \textit{other factors}$$

… but this assumes the ratings equal-interval property is satisfied.

Option B. regress municipal bond ratings on four dummies:

$$MBR = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + \textit{other factors}.$$

If there were data available, then one could use an F test to compare model fit with 1 variable with k values (Option A) vs. k–1 dummies (Option B); refer back to Lecture on 'ANOVA for Regression Specifications'

2. Multiple Dummies for a Single Dimension in a Regression

Consider an *ordinal* explanatory variable…

e.g., how law students' starting salaries depending on rankings [Example 7.8, 239–40]:

$$\widehat{\log(salary)} = \underset{(.41)}{9.17} + \underset{(.053)}{.700}\, top10 + \underset{(.039)}{.594}\, r11\_25 + \underset{(.034)}{.375}\, r26\_40 + \underset{(.028)}{.263}\, r41\_60 + \underset{(.021)}{.132}\, r61\_100 + \underset{(.0031)}{.0057}\, LSAT$$

$n = 136$, $R^2 = .911$,

Use dataset lawsch85.dta and see R script: Lecture 23A law students

Use an F test to compare fit using 1 variable with 6 values vs. 5 dummies
$df_r = 133 = 136 - 3$ (constant, *LSAT*, one ordinal variable)
$df_u = 129 = 136 - 7$ (constant, *LSAT*, five binary variables)
$q = df_r - df_u = 4$

$$F = \frac{\left(R_u{}^2 - R_r{}^2\right)/4}{\left(1 - R_u{}^2\right)/129}$$

Don't forget the rule for interpreting marginal effects in a log-linear model: the *percentage* change in salary for group j, relative to baseline group, $= 100 \cdot \left(e^{\delta_j} - 1\right)$

3. Hypothesis Testing for Unidimensional Dummies – Review [4.2]

Null hypothesis: $H_{0j} : \beta_j = 0$

Alternative hypothesis: $H_{1j} : \beta_j \neq 0$

Test statistic: $t = \dfrac{\widehat{\beta_j}}{se(\widehat{\beta_j})}$

The t statistic answers the question: How many standard errors separate the estimated coefficient ($\widehat{\beta}_j$) from the null hypothesis (0) for the coefficient?

Values of the t statistic sufficiently far from 0 result in a rejection of H_{0j} ;
our **rejection rule** is to reject H_{0j} if and only if $\left|t_{\widehat{\beta_j}}\right| > c$;
by convention, c is chosen to make areas above $+c$ and below $-c$ sum to .05

If H_{0j} is rejected at the 5% level, then we conventionally say "x_j is statistically significant."

POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

3. Hypothesis Testing for Unidimensional Dummies – Review [4.2]

Null hypothesis: $H_{0j} : \beta_j = a_j$

Alternative hypothesis: $H_{1j} : \beta_j \neq a_j$

Test statistic: $$t = \frac{(estimate - hypothesized\ value)}{standard\ error\ of\ estimate} = \frac{(\widehat{\beta_j} - a_j)}{se(\widehat{\beta_j})}$$

The t statistic answers the question: How many standard errors separate the estimated coefficient ($\widehat{\beta_j}$) from the null hypothesis (a_j) for the coefficient?

Values of the t statistic sufficiently far from 0 result in a rejection of H_{0j} ;
our **rejection rule** is to reject H_{0j} if and only if $\left|t_{\widehat{\beta_j}}\right| > c$;
by convention, c is chosen to make areas above $+c$ and below $-c$ sum to .05

If H_{0j} is rejected at the 5% level, then we conventionally say "x_j is statistically significant."

3. Hypothesis Testing for Unidimensional Dummies [4.4]

Null hypothesis: $H_0 : \beta_1 - \beta_2 = 0$ (equivalent to $H_0 : \beta_1 = \beta_2$)

Alternative hypothesis: $H_1 : \beta_1 - \beta_2 \neq 0$ (equivalent to $H_1 : \beta_1 \neq \beta_2$)

Test statistic is $t = \dfrac{difference\ of\ coefficients}{standard\ error\ of\ difference\ \ldots} = \dfrac{\widehat{\beta_1} - \widehat{\beta_2}}{se\left(\widehat{\beta_1} - \widehat{\beta_2}\right)}$

where $se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{\widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) \boxed{- 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}}$

$$= \sqrt{\left[\widehat{Var}(\hat{\beta}_1) - \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)\right] + \left[\widehat{Var}(\hat{\beta}_2) - \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)\right]}$$

Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

4. Example: *'Looking Good on Course Evaluations' part 2*

http://chance.amstat.org/2013/04/looking-good/
Data: eval.csv
Script: Lecture 23A evaluations.R

Dependent variable is average evaluation of professor, prof_eval (0 – 5 scale; mean = 4.175)

Explanatory variable is average attractiveness bty_avg (range = 1.667 – 8.167; mean = 4.418)

"Level one" control variables:
age (quantitative)
gender (binary; *female* or *male*)
language (binary; *English* or *non-English*)
ethnicity (binary; *minority* or *not minority*)
rank (qualitative; *teaching* and *tenure track* and *tenured*)

"Level two" variables:
six different raters (three male, three female)
thirty classes (class1, …, class30)
cls_level (binary; *lower* division or upper division)

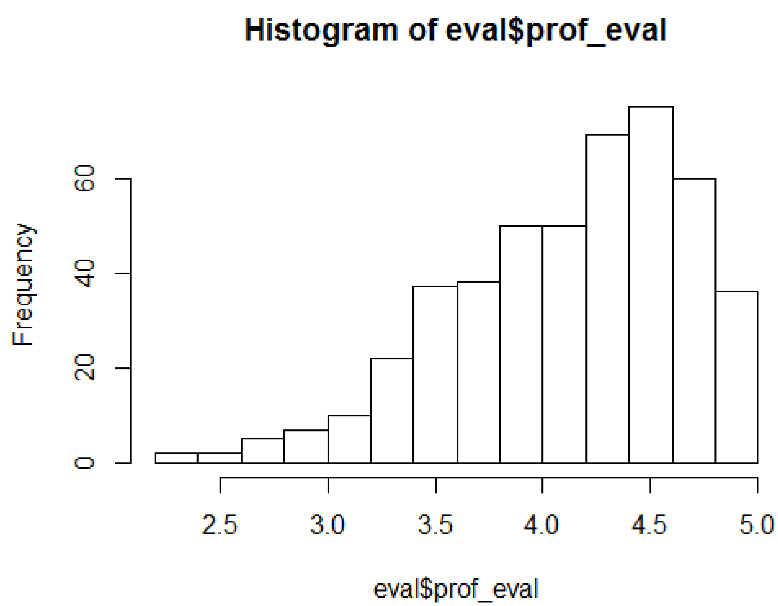
4. Example: *'Looking Good on Course Evaluations' part 2*

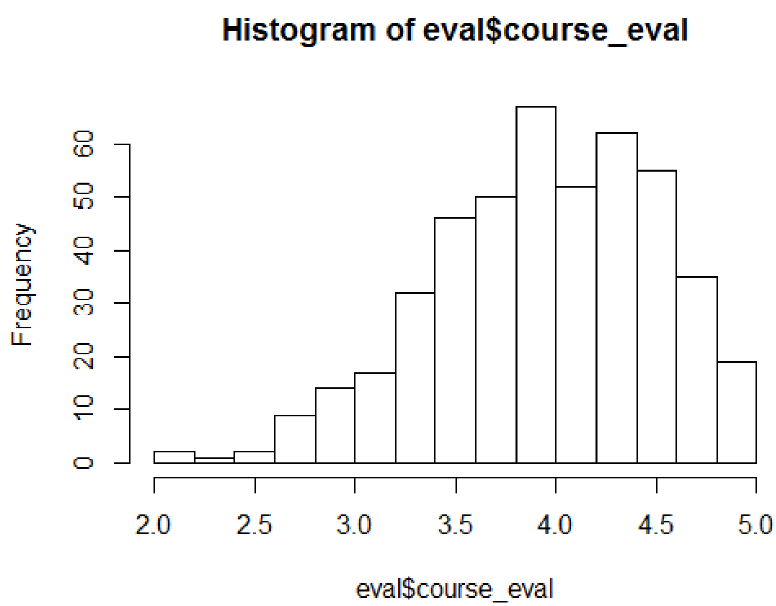
http://chance.amstat.org/2013/04/looking-good/

Data: eval.csv

Script: Lecture 23A evaluations.R

Dependent variable is average evaluation of professor (0 – 5 scale; mean = 4.175)
[average evaluation of course (0 – 5 scale; mean = 3.998)]





POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 7.1 + 7.2a + 7.3 (+ 4.2 + 4.4) + 17.3

4. Example: *'Looking Good on Course Evaluations' part 2*

Consider the effects of rank ("teaching" and "tenure track" and "tenured")
Dependent variable is average evaluation of professor (0 – 5 scale; mean = 4.175)

Without control for attractiveness (R^2 = .0116):

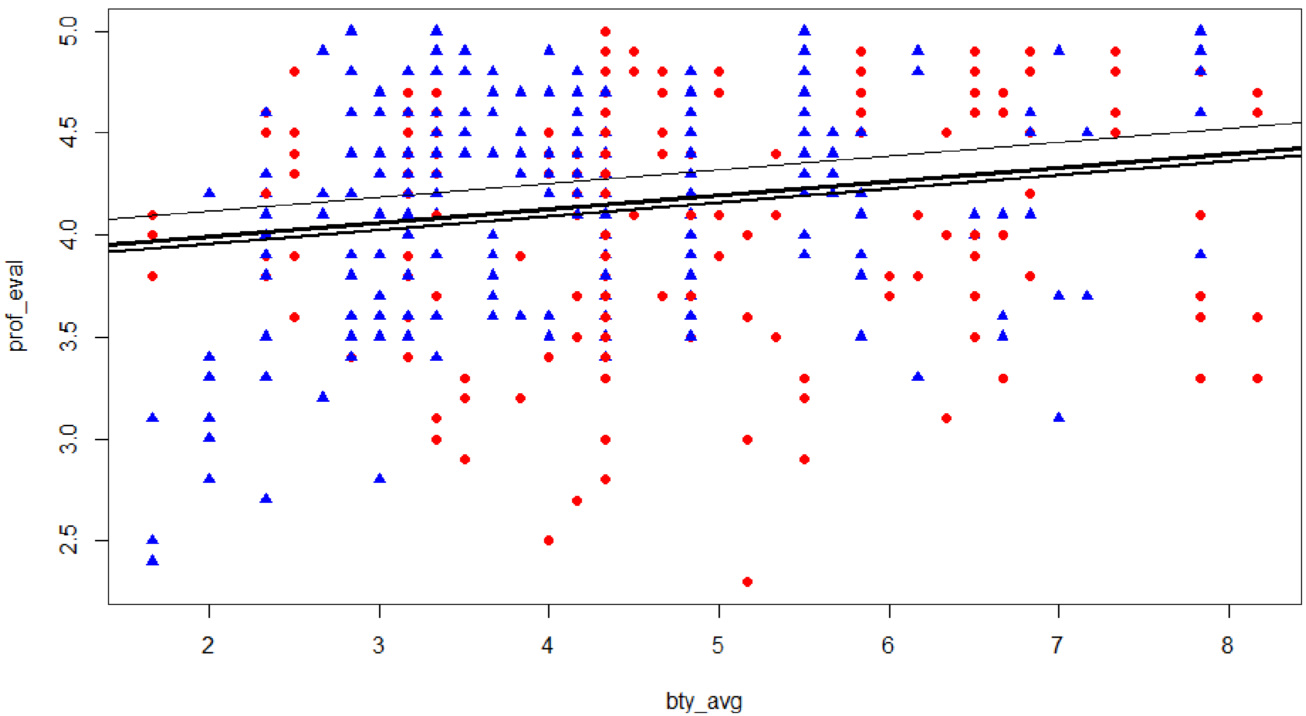
$$\underset{}{\widehat{prof\_eval}} = \underset{(.0537)}{4.2843} - \underset{(.0636)}{.1452}\ tenured - \underset{(.0748)}{.1297}\ tenure\ track$$

With a control for attractiveness (R^2 = .0465):

$$\widehat{prof\_eval} = \underset{(.0908)}{3.9816} - \underset{(.0627)}{.1262}\ tenured - \underset{(.0740)}{.1607}\ tenure\ track + \underset{(.0166)}{.0678}\ bty\_avg$$

4. Example: *'Looking Good on Course Evaluations' part 2*

Consider the effects of rank ("teaching" and "tenure track" and "tenured")
Dependent variable is average evaluation of professor (0 – 5 scale; mean = 4.175)



4. Example: *'Looking Good on Course Evaluations' part 2*

Consider the effects of rank ("teaching" and "tenure track" and "tenured")

Null hypotheses: $H_{0\text{A}} : \beta_1 = 0;$ $H_{0\text{B}} : \beta_2 = 0$

Alternative hypotheses: $H_{1\text{A}} : \beta_1 \neq 0;$ $H_{1\text{B}} : \beta_2 \neq 0$

Test statistics are $t_A = \dfrac{\widehat{\beta_1} - 0}{se(\widehat{\beta_1})};$ $t_B = \dfrac{\widehat{\beta_2} - 0}{se(\widehat{\beta_2})}$

$$\text{Estimated } \widehat{prof\_eval} = \underset{(.0908)}{3.9816} - \underset{(.0627)}{.1262}\ tenured - \underset{(.0740)}{.1607}\ tenure\ track + \underset{(.0166)}{.0678}\ bty\_avg$$

t statistics: $t_{tenured} = \dfrac{.1262}{.0627} = 2.013$ $t_{tenure\ track} = \dfrac{.1607}{.0740} = 2.172$

4. Example: *'Looking Good on Course Evaluations' part 2*

Reconsider the effects of rank ("teaching" and "tenure track" and "tenured")

Test statistic for difference between *tenured* and *tenure track*: $t = \dfrac{\widehat{\beta_1} - \widehat{\beta_2}}{se(\widehat{\beta_1} - \widehat{\beta_2})}$

Estimated $\widehat{prof\_eval}$ = 3.9816 – .1262 *tenured* – .1607 *tenure track* + .0678 *bty_avg*
(.0908) (.0627) (.0740) (.0166)

Numerator: $\widehat{\beta_1} - \widehat{\beta_2}$ = (– .1262) – (–.1607) = .0345

Denominator: $se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{\widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) - 2\boxed{\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}}$

For standard errors, need variance-covariance matrix of estimators:

	track	tenure
track	0.00547	0.00275
tenure	0.00275	0.00393

standard error equals √.0039 = .06245; *t* statistic equals ±.552; *F* statistic = .3048
Retain null hypothesis of no difference.

5A. Review of Poisson Probability Distribution

$$P(y = h) = e^{-\mu}\frac{\mu^h}{h!}, h = 0, 1, 2, \ldots$$

Flying bomb attacks during Second World War; London divided into squares, each .5 km. wide and .5 km. tall

Bombs	Observed Squares	Predicted Squares
0	229	
1	211	
2	93	
3	35	
4	7	
5	1	

Make predictions based on $\mu = 0.9288 = 535$ bombs $\div$ 576 squares

Code for frequencies: `576*dpois(c(0:5), .9288)`
Code for cumulative frequencies: `576*ppois(c(0:4), .9288, lower.tail = TRUE)`

R.D. Clark (1946) "An Application of the Poisson Process" Journal of the Institute of Actuaries 72: 481–

5B. Review of Poisson Regression

$$P(y = h) = e^{-\mu}\frac{\mu^h}{h!}, h = 0, 1, 2, \ldots$$

Model μ as: $\mu(x) = \exp(x\hat{\beta}) = e^{\widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \ldots}$

Combine: $P(y = h) = e^{-\exp(x\hat{\beta})}\dfrac{\exp(x\hat{\beta})^h}{h!}$

Predict: set fixed values of x_1, x_2, … , and then change them one at a time

Interpret $\hat{\beta}_j$: $100\cdot[\exp(\hat{\beta}_j\,\Delta x_j)-1]$ tells you the percent or proportional change in $\mu(x)$ from a Δx_j–unit change in x_j

5C. Review of Poisson Regression

Application: *Traffic Accidents at Intersections*

Let y_i denote the number of incidents in an intersection in a given year
Let $x_{1i} \in \{25, 30, 35, \ldots\}$ denote the average speed on nearby streets, in mph
Let $x_{2i} \in \{0, 1\}$ denote whether intersection has a traffic signal

Suppose : $\widehat{\beta_0} = 2.8, \widehat{\beta_1} = 0.012, \widehat{\beta_2} = -0.20$

Model μ as : $\mu(x) = \exp(x\hat{\beta}) = e^{\widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2} = e^{2.8 + .012x_1 - .2x_2}$

Combine : $P(y = h) = e^{-\exp(x\hat{\beta})} \dfrac{\exp(x\hat{\beta})^h}{h!}$

Predict : set $x_1 = 25$, set $x_2 = 0$ and predict, then reset $x_1 = 35$ and/or reset $x_2 = 1$

R code:

```
mu.25.0 = exp(2.8 + .012*25); round(dpois(c(17:26), mu.25.0), digits = 2)

mu.25.1 = exp(2.6 + .012*25); round(dpois(c(17:26), mu.25.1), digits=2)

mu.35.0 = exp(2.8 + .012*35); round(dpois(c(17:26), mu.35.0), digits=3)

mu.35.1 = exp(2.6 + .012*35); round(dpois(c(17:26), mu.35.1), digits=3)
```

5C. Review of Poisson Regression

Application: *Traffic Accidents at Intersections*

Let y_i denote the number of incidents in an intersection in a given year
Let $x_{1i} \in \{25, 30, 35, \ldots\}$ denote the average speed on nearby streets, in mph
Let $x_{2i} \in \{0, 1\}$ denote whether intersection has a traffic signal

Suppose : $\widehat{\beta_0} = 2.8, \widehat{\beta_1} = 0.012, \widehat{\beta_2} = -0.20$

Model μ as : $\mu(x) = \exp(x\hat{\beta}) = e^{\widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2} = e^{2.8 + .012x_1 - .2x_2}$

Combine : $P(y = h) = e^{-\exp(x\hat{\beta})} \dfrac{\exp(x\hat{\beta})^h}{h!}$

Predict : set $x_1 = 25$, set $x_2 = 0$ and predict, then reset $x_1 = 35$ and/or reset $x_2 = 1$

Interpret $\hat{\beta}_1$: $100 \cdot [\exp(\hat{\beta}_1 \cdot 10) - 1]$ tells you the percent or proportional change in $\mu(x)$ from increasing x_1 by 10 mph.

R code: `(mu.25.1/mu.25.0)-1; (mu.35.1/mu.35.0)-1`

Interpret $\hat{\beta}_2$: $100 \cdot [\exp(\hat{\beta}_2) - 1]$ tells you the percent or proportional change in $\mu(x)$ from adding a traffic signal.

R code: `(mu.35.0/mu.25.0)-1; (mu.35.1/mu.25.1)-1`

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Example: *Economic Model of Crime*

Dataset is crime1.RData; Wooldridge uses in examples 7.12, 8.3, 9.1, and 17.3

Population is 'young men in California born in 1960 or 1961 who have at least one arrest prior to 1986'; sample is 2,725 individuals

Dependent variable: *narr86* (number of arrests in 1986; 27.7% arrested ≥ 1x in 1986, however only 7.2% were arrested > 1x)

Explanatory variables:
pcnv (proportion of prior arrests leading to a conviction)*
avgsen (average sentence served from prior convictions, in months)*
tottime (time spent in prison since age 18 up to 1985, in months)
ptime86 (time spent in prison during 1986, in months)*
qemp86 (quarters employed during 1986, 0 – 4)
inc86 (legal income, in hundreds of $)*
black (dummy variable)
hispanic (dummy variable)
born60 (dummy variable)

* squared terms added to specification in various examples

POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Example: *Economic Model of Crime*

Dataset is crime1.RData; Wooldridge uses in examples 7.12, 8.3, 9.1, and 17.3

TABLE 17.5 Determinants of Number of Arrests for Young Men

Dependent Variable: *narr86*		
Independent Variables	Linear (OLS)	Exponential (Poisson QMLE)
pcnv	−.132 (.040)	−.402 (.085)
avgsen	−.011 (.012)	−.024 (.020)
tottime	.012 (.009)	.024 (.015)
ptime86	−.041 (.009)	−.099 (.021)
qemp86	−.051 (.014)	−.038 (.029)
inc86	−.0015 (.0003)	−.0081 (.0010)
black	.327 (.045)	.661 (.074)
hispan	.194 (.040)	.500 (.074)
born60	−.022 (.033)	−.051 (.064)
constant	.577 (.038)	−.600 (.067)

6. Example: *Economic Model of Crime*

see R script: Lecture 23B crime.R

Example:
```
load("C:/crime1.RData"); attach(data)
pois.D <- glm(narr86 ~ pcnv + ptime86 + qemp86 + black + hispan, poisson)
reg.D <- lm(narr86 ~ pcnv + ptime86 + qemp86 + black + hispan)
```

Compare:
```
stargazer(reg.D, pois.D, type="text", single.row=FALSE, omit.stat=c("f", "ser"))
```

Interpret: for each quarter that man is employed, arrests decrease by 20% *proportionately*
black men are arrested 103% more often than white men
hispanic men are arrested 70% more often than white men

6. Example: *Economic Model of Crime*

Recall from earlier : $P(y = h) = e^{-\exp(x\hat{\beta})} \dfrac{\exp(x\hat{\beta})^h}{h!}$

where : $\exp(x\hat{\beta}) = e^{\widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \ldots}$

Use $\widehat{\beta_0} = -.578$, $\hat{\beta}_{pcnv} = -.379$, $\hat{\beta}_{ptime86} = -.096$, $\hat{\beta}_{qemp86} = -.226$, $\hat{\beta}_{black} = .710$, $\hat{\beta}_{hispan} = .533$

Suppose a male has no convictions (*pcnv* = 0), spent no time in prison in 1986 (*ptime86* = 0), and was employed for 2 quarters in 1986 (*qemp86* = 2).

If that man is white (*black* = 0, *hispan* = 0), then:	$x\beta = -1.030$;	$\exp(x\beta) = .357$
If that man is black (*black* = 1, *hispan* = 0), then:	$x\beta = -0.320$;	$\exp(x\beta) = .726$
If that man is hispanic (*black* = 0, *hispan* = 1), then:	$x\beta = -0.498$;	$\exp(x\beta) = .608$

Probabilities of no arrests (h = 0):	white:	70.0 %
	black:	48.4 %
	hispanic:	54.4 %

Probabilities of one arrest (h = 1):	white:	25.0 %
	black:	35.1 %
	hispanic:	33.1 %

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Example: *Economic Model of Crime*

see R script: Lecture 23B crime.R

Example:

```
load("C:/crime1.RData"); attach(data)
pois.D <- glm(narr86 ~ pcnv + inc86 + black + hispan, poisson)
```

Dependent variable *(narr86)* is number of arrests in 1986

Model with controls for prior convictions and for income in 1986 (R^2 = .069):

$$\widehat{narr86} = \underset{(.057)}{-.676} \; - \underset{(.085)}{.433}\, pcnv \; - \underset{(.001)}{.009}\, inc86 \; + \underset{(.074)}{.471}\, hispan \; + \underset{(.073)}{.644}\, black$$

A one-unit ($100) increase in *inc86* changes expected number of arrests by –0.9%

Interpret $\hat{\beta}_{inc86}$: $100\cdot[\exp(\hat{\beta}_{inc86}\cdot 1)-1]$ tells you the percent or proportional change in $\mu(x)$ from increasing *inc86* by $100

Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Example: *Economic Model of Crime*

see R script: Lecture 23B crime.R

Example:
```
load("C:/crime1.RData"); attach(data)
pois.D <- glm(narr86 ~ pcnv + inc86 + black + hispan, poisson)
```

Dependent variable *(narr86)* is number of arrests in 1986

Model with controls for prior convictions and for income in 1986 (R^2 = .069):

$$\widehat{narr86} = \underset{(.057)}{-.676} \; \underset{(.085)}{-\, .433}\, pcnv \; \underset{(.001)}{-\, .009}\, inc86 \; \underset{(.074)}{+\, .471}\, hispan \; \underset{(.073)}{+\, .644}\, black$$

A one-unit 'increase' in *hispan* changes expected number of arrests by +60%

Interpret $\hat{\beta}_{hispan}$: $100 \cdot [\exp(\hat{\beta}_{hispan} \cdot 1) - 1]$ tells you the percent or proportional change in $\mu(x)$ from changing *hispan* from 0 to 1

6. Example: *Economic Model of Crime*

see R script: Lecture 23B crime.R

Example:

```
load("C:/crime1.RData"); attach(data)
pois.D <- glm(narr86 ~ pcnv + inc86 + black + hispan, poisson)
```

Dependent variable *(narr86)* is number of arrests in 1986

Model with controls for prior convictions and for income in 1986 (R^2 = .069):

$$\widehat{narr86} = \underset{(.057)}{-.676} - \underset{(.085)}{.433}\, pcnv - \underset{(.001)}{.009}\, inc86 + \underset{(.074)}{.471}\, hispan + \underset{(.073)}{.644}\, black$$

A one-unit 'increase' in *black* changes expected number of arrests by +90%

Interpret $\hat{\beta}_{black}$: $100 \cdot [\exp(\hat{\beta}_{black} \cdot 1) - 1]$ tells you the percent or proportional change in $\mu(x)$ from changing *black* from 0 to 1

POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Example: *Economic Model of Crime*

Recall from earlier : $P(y = h) = e^{-\exp(x\hat{\beta})} \dfrac{\exp(x\hat{\beta})^h}{h!}$

where : $\exp(x\hat{\beta}) = e^{\widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \ldots}$

Use $\widehat{\beta_0} = -.676$, $\hat{\beta}_{pcnv} = -.433$, $\hat{\beta}_{inc86} = -.009$, $\boxed{\hat{\beta}_{hispan} = .471\ , \hat{\beta}_{black} = .644}$

Suppose a man has no convictions ($pcnv = 0$) and earned \$5,500 income in 1986 ($inc86 = 55$)

				Probabilities of:	
r/e:	$x\beta$	$\exp(x\beta)$	no arrests (h = 0):	one arrest (h = 1):	two arrests (h = 2)
white	–1.152	= .316	72.9 %	23.0 %	3.6 %
Hispanic	–0.681	= .506	60.3 %	30.5 %	7.7 %
black	–0.508	= .602	54.8 %	33.0%	9.9 %

6. Example: *Economic Model of Crime*

Consider the effects of race/ethnicity (white vs. Hispanic, white vs. black)

Dependent variable *(narr86)* is number of arrests in 1986

Model with controls for prior convictions and for income in 1986 (R^2 = .069):

$$\widehat{narr86} = -.676 - .433\, pcnv - .009\, inc86 + .471\, hispan + .644\, black$$

Std. errors:	(.057)	(.085)	(.001)	(.074)	(.073)
Robust std. errors:	(.074)	(.103)	(.001)	(.094)	(.099)

Initial tests of hypothesis:

$$\mathrm{H}_{0h} : \hat{\beta}_{hispan} = 0 \qquad z = \frac{.471}{.074} = 6.403 > 1.96 \qquad z = \frac{.471}{.094} = 5.041 > 1.96$$

$$\mathrm{H}_{0b} : \hat{\beta}_{black} = 0 \qquad z = \frac{.644}{.073} = 8.777 > 1.96 \qquad z = \frac{.644}{.099} = 6.499 > 1.96$$

POLS 6481. Research Design and Quantitative Methods II
Lecture 23. Binary Independent Variables I: Different Intercepts for Different Groups
Readings: Wooldridge, *Introductory Econometrics 5e*, 17.3

6. Example: *Economic Model of Crime*

Consider the effects of race/ethnicity (Hispanic vs. black)

The null hypothesis is: $H_0 : \beta_1 = \beta_2$

The test statistic is: $z = \frac{\widehat{\beta_1} - \widehat{\beta_2}}{se\left(\widehat{\beta_1} - \widehat{\beta_2}\right)} = \frac{difference\ of\ coefficients}{standard\ error\ of\ difference\ of\ coefficients}$

where $se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{\widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) - 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$

Numerator: $\hat{\beta}_{black} - \hat{\beta}_{hispanic} = .644 - .471 = .173$

For standard error, we need variance-covariance matrix of estimators:

	black	hispanic
black	0.005383	0.001989
hispanic	0.001989	0.005422

Denominator: $\sqrt{.005383 + .005422 - 2 \cdot .001989} = \sqrt{.006828} = .0826$

z statistic equals 2.087 > 1.96

6. Example: *Economic Model of Crime*

Consider the effects of race/ethnicity (Hispanic vs. black)

The null hypothesis is: $H_0 : \beta_1 = \beta_2$

The test statistic is: $z = \frac{\widehat{\beta_1} - \widehat{\beta_2}}{se\left(\widehat{\beta_1} - \widehat{\beta_2}\right)} = \frac{difference\ of\ coefficients}{standard\ error\ of\ difference\ of\ coefficients}$

where $se(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{\widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) - 2\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2)}$

Numerator: $\hat{\beta}_{black} - \hat{\beta}_{hispanic} = .644 - .471 = .173$

For robust standard errors, we need new variance-covariance matrix of estimators:

	black	hispanic
black	0.009816	0.003243
hispanic	0.003243	0.008746

Denominator: $\sqrt{.009816 + .008746 - 2 \cdot .003243} = \sqrt{.006828} = .1099$

z statistic equals 1.5695 < 1.96