Concept	Applicable Models	Formula	Interpretation
t-test	Linear	$t_{\widehat{\beta}} = \dfrac{\widehat{\beta} - \beta_0}{\mathrm{SE}(\widehat{\beta})}$	Observed data is t number of standard deviations away from mean, also sqrt(F) = t if one coef.
z-test	Logistic	$z = \dfrac{\hat{\beta}_j}{\hat{se}(\hat{\beta}_j)}$	Observed data is Z number of standard deviations away from mean
Wald Test	Logistic	$W = \left(\dfrac{\hat{\beta}_j}{\hat{se}(\hat{\beta}_j)}\right)^2$	.Can test both single coefficients and entire models.
F-test	Linear	$\dfrac{SSM/k}{SSE/[n-(k+1)]}$ OR $F = \dfrac{R^2/k}{(1-R^2)/[n-(k+1)]}$ OR $\dfrac{R^2_{\text{change}}/df_{\text{change}}}{(1-R_a^2)/[n-(k_a+1)]}$	1 or more explanatory variables in a multiple linear regression are (not) significant. K is change in df. R^2a & ka is for full model. Measures the ratio of explained variance vs. unexplained variance in dependent varaible.
Likelihood Ratio Test	Logistic	$2\log(L_2) - 2\log(L_1)$	Expanded model is(not) a better explainer of outcome than nested model. *At least one of the new coefficients are not 0*.

Computing Probabilities	
Bi Log	$\pi = \dfrac{\exp(\text{logit})}{1+\exp(\text{logit})} = \dfrac{\text{Odds}}{1+\text{Odds}}$ $\text{logit} = \log\left(\dfrac{\pi_i}{1-\pi_i}\right) = \alpha$
Multi (ref)	1 / bottom half of non-ref
Multi (non-ref)	$\hat{\pi}_j = \dfrac{\exp[\hat{\alpha}_j + \hat{\beta}_{j1}X_1 + \hat{\beta}_{j2}X_2 + \cdots + \hat{\beta}_{jk}X_k]}{1+\sum_{l=1}^{c-1}\exp[\hat{\alpha}_l + \hat{\beta}_{l1}X_1 + \hat{\beta}_{l2}X_2 + \cdots + \hat{\beta}_{lk}X_k]}$
Ordinal	$P(Y \le j) = \dfrac{\exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1+\exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}$
Poisson	$p(y\mid\lambda) = \dfrac{\lambda^y e^{-\lambda}}{y!}$ $E[y\mid\lambda] = \lambda$ $Var[y\mid\lambda] = \lambda$
Negative Binomial	E[y I lambda] = lambda Var[y] = lambda + alpha*lambda^2

Where we are in the course

Linear regression model:

$$E(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Binary logistic models:

$$\log\left[\frac{P(Y=1)}{P(Y=0)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Multinomial logistic models:

$$\log\left[\frac{P(Y=j)}{P(Y=0)}\right] = \alpha_j + \beta_{j1} X_1 + \beta_{j2} X_2 + \cdots + \beta_{jk} X_k$$

Ordinal logistic models:

$$\log\left[\frac{P(Y\le j)}{P(Y>j)}\right] = \alpha_j - (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

Count models:

$$\log[E(Y)] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

Model Parameters		
Concept	**Formula**	**Definition**
TSS/SS (Total Sum of Squares)	TSS = SSE + SSE; TSS = (Yi - mean(Y))^2	*measure of the total variation in Yi in the sample. ; TSS/(n-1) = sample variance of y.*
SSE/RSS (Sum Squared Error or Residual Sum of Squares)	SSE = sum((Yi – Y-hat)^2); SSE = TSS-SSM	*total unexplained variation in our model. Parameter estimation that minimises this is the least squares estimate.*

SSM (Model Sum of Squares)	SSM = sum((Yi-hat – Y-mean)^2); SSM = TSS-SSE	*The reduction in squared prediction errors achieved when we make use of Xi to predict values of Yi instead of just using the mean of Y.*
Residual SD (RSE)	$\hat{\sigma} = \sqrt{(SSE / (n - p - 1))}$	*SD of all the residuals from our prediction vs. observed data.*
R-squared (Coef of determ)	R^2 = (TSS-SSE)/TSS; 1 – SSE/TSS; 1 – sum((Yi – Y-hat)^2)/sum((Yi – Y-mean)^2)	*measure of the proportion of total variation of Y in the sample explained by our model as proportion of TSS. ; sqrt of R^2 is correlation between Yi and Y-hat-I values.*
Adjusted R-squared	$1 - \left[\frac{(1 - R^2) \times (n - 1)}{(n - k - 1)}\right]$	*R^2 adjusted for number of expl parameters to address limitation.*
Pseudo-R^2 (MacFaddens)	$\text{Pseudo}-R^2 = \frac{-\log L_N - (-\log L_1)}{-\log L_N}$	*Generalising R^2 from linear to logistic through measuring relative to null model (intercept-only).*
Akaike's Info Criterion (AIC)	-2log(likelihood) + 2k K includes intercept here.	*Measures that penalise the log-likelihood by some measure of model complexity (k) (smaller the better).*

Properties of logarithms and exponential functions

- $\log(e) = 1$
- $\log(1) = 0$
- $\log(x^r) = r \log(x)$
- $\log(e^A) = A$
- $e^{\log A} = A$
- $\log(AB) = \log A + \log B$
- $\log(A/B) = \log A - \log B$
- $e^{AB} = (e^A)^B$
- $e^{A+B} = e^A e^B$
- $e^{A-B} = e^A / e^B$

Note: Here all logarithms are "natural" logs, that is, log to base e, sometimes denoted "ln"

$$\begin{aligned}
\hat{Y} &= \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k \\
\hat{\pi} &= \frac{1}{1 + \exp[-(\hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k)]} \\
\hat{\pi}_j &= \frac{\exp[\hat{\alpha}_j + \hat{\beta}_{j1} X_1 + \hat{\beta}_{j2} X_2 + \cdots + \hat{\beta}_{jk} X_k]}{1 + \sum_{l=1}^{c-1} \exp[\hat{\alpha}_l + \hat{\beta}_{l1} X_1 + \hat{\beta}_{l2} X_2 + \cdots + \hat{\beta}_{lk} X_k]} \\
\gamma^{(j)} &= P(Y \leq j) = \frac{\exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]}{1 + \exp[\alpha^{(j)} - (\beta_1 X_1 + \cdots + \beta_k X_k)]} \\
\hat{\lambda} &= \exp(\hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 \ldots + \hat{\beta}_k X_k
\end{aligned}$$

2^3 = 8; b^p = n; log(base-b)n = p; log(base-2)8 = 3. How many times do I multiply b by itself to make n = p. log(base-10)10,000 = 4 – 10 raised to what power = 10,000? 4. To get natural log from log10, multiply by 2.303. *A change from 2 years to 20 years of education is equivalent to an increase in the log10(educ) by 1 unit. Thus, 1 unit increase in the x-variable, log10(educ), is associated with a change in the log of surveys taken by 0.185. We could also say 10 times increase in the years of education (which is a 1 unit increase in the log10) multiples the expected number of surveys taken by exp(0.185) = 1.203.* For 5c I treated it as a question of "what effect does an increase in age of 18 years have on the expected number of surveys"... education in the model is base10 log(educ), the diff in logged education is log(20) - log(2) = log(20/2) = 1.

Odds & Odds Ratios

Odds = prob-i / 1 – prob-i

- If the probability of being a victim of crime is pi = 0.2107

- The odds of being a victim of crime are 0.2107/0.7893 = 0.2669
- The odds of not being a victim of crime are 0.7893/0.2107 = 1/0.2669/ 3.7461
- Remember to include interactions when computing probabilities.

Model Assumptions

<u>Linear Regression</u>

Assumption 1: Linearity

1. *Definition: linear (additive) relationship between dependent variable and each independent variable.*

Assumption 2: Independence of Observations

1. *Definition: Observations are independent of each other and don't affect one another.*

Assumption 3: Homoscedasticity

1. *Definition: Variance of the errors is constant across all levels of the independent variable.*

Assumption 4: Normality

1. *Definition: errors of the model should be normally distributed.*

Assumption 5: No multicollinearity

1. *Definition: No perfect or high correlation between independent variables.*

Assumption 6: sufficient sample size

<u>Multinomial Logistic Regression</u>

Assumption 1: Independence of Irrelevant Alternatives

1. *Definition: a condition that states that the relative likelihood of choosing from A from B won't change if a third choice is placed into the mix.*
2. *Definition 2: IIA implies that the presence or absence of an alternative has no effect on the relative proportion of individuals choosing among the remaining alternatives.*
3. *Result: We can't say anything about 'counterfactual' that is not observed.*

Assumption 2: Linearity

Assumption 3: Independence of Observations

Assumption 4: no multicollinearity

Assumption 5: Multinomial distribution

1. *Definition: dependent variable should follow multinom distr – have 2 or more categories and be mutually exclusive and exhaustive.*

Assumption 5: sufficient sample size

Assumption 6: DFs are determined by the number of outcome categories multiplied by the number of new coefs per

<u>Ordinal Logistic Regression</u>

Assumption 1: Proportional Odds Assumption

1. *Definition: Curves for values of Y are 'parallel,' meaning that all coefficients (except alpha) have consistent values for all values of Y.*
    a. *This implies that odds are not conditional on values of covariates.*

Assumption 2: Ordinality of Outcomes

1. *Definition: outcome variable is naturally ordered.*
    a. *Differs from logistic models which assume linearity rather than ordinality.*

Assumption 3: Intercept represents the distance between curves and threshold value that would tip observation from being in lower to higher category. Can interpret with 'setting all covariates in kodel to zero' odds of..

<u>Poisson Count</u>

Assumption 1: Homogeneity of Variance

1. *Definition: variance of count variable is constant across all levels of explanatory vars*

Assumption 2: No overdispersion

1. *Definition: variance equal to mean*

Additionals: Independence, Count data, Poisson distribution, linearity, multicollinearity, sufficient sample size

Template Responses

Interpreting Coefficients

1. Are we talking about a *bivariate* (single-coefficient), *partial* (holding all other expl vars constant) or a *partial relationship conditional* on the relationship between X2? 1/Multiple/Interactions (cant hold constant all other expl variables), Interaction changes slopes.
2. What is outcome space? Change in what are we modelling here?

- We are 95% confident that the population parameter lies between LB & UB.

Hypothesis Testing

Test (sampling distr.)	Number of Parameters Tested 1	2+
Linear Regression	*t*-test (t)	*F*-test (F)
Maximum Likelihood	*z*-test ($\mathcal{N}$) Wald test (χ^2)	Likelihood Ratio test (χ^2)

- Two types of model: one with a single coefficient not equalling 0 and another where *at least one* is not 0. So you're testing whether **all** coefficients = 0 in second case.
- Significance levels: 0.1 = 1.645, 0.05 = 1.96, 0.01 = 2.576.
- P-values, generally, can indicate how incompatible the data are with a specified statistical model.
- Expect value at least as large/extreme as the one observed. SS at level 0.0X and thus can/not reject the null hypothesis that Beta-X = 0.
- Bonferroni correction (inflated risk of false positive when increasing parameters) - divides p-value by number of parameters, which is new p-value to aim for.

Linear

- Degrees of freedom = n – k - 1
- increasing Xk by 1 unit while holding other Xs constant increases expected value of Y by βk units

Binary Log

- increasing Xk by 1 unit while holding other Xs constant multiplies the odds of Y= 1 by exp(βk).

Multi Log

- Increasing Xk by 1 unit while holding other Xs constant multiplies the odds of Y having the value j rather than the reference value 0 by exp(βjk).
    - 0.972 = exp(Beta-age): On average, controlling for whether or not a person has been stopped, a 1-year increase in age multiplies the odds of having been a victim of crime by 0.972. i.e, decrease the odds by 2.8%.
- Holding 'controlling covariates' constant, the odds for someone with 'Conditional term' of voting Y1 rather than Y0 are ???? times (???% higher than) the odds for someone with 'alternative conditional term' or less.
    - 1.592 = exp(Beta-stop): On av, controlling for other expl vars, having been stopped (stop = 1) multiplies the odds of having been a victim of crime by 1.592, compared to not having been stopped (stop = 0), i.e., increases the odds by 59.2%.

Ordinal

- Exponentiated coefficient estimates are interpreted as partial odds ratios for being in the higher rather than the lower half of the dichotomy.
    - e.g., exp(β^male) = 0.48: Controlling for the other explanatory variables, (we estimate that) men have 52% lower odds than women of giving a response that indicates higher levels of agreement with the statement "XYZ"
    - e.g., exp(β^ ed) = 1.088: Controlling for the other explanatory variables, 1 additional year of education is associated with an 8.8% increase in (estimated) odds of giving a response that indicates higher levels of agreement with the statement
    - exp(β^ ed ± 1.96 ∗ se(β^ ed)) = [1.06; 1.12]: with 95% confidence we estimate that [...] 1 additional year of education increases the odds... 95% confidence, not 95% probability, as the true parameter is either in it or isn't.
    - alpha/intercept: setting all covariates in the model to zero the odds of saying politics is not_important rather than [alternative response is [interpret coef as usual].

Poisson

- increasing Xk by 1 multiplies E[Y|X] by e^Beta-k
- Round to integer if relevant.
- increasing Xk by 1 unit while holding other Xs constant multiples the expected value of Y by eβk.